IDENTIFICATION OF THE GENETIC AND PHENOTYPIC BASIS OF ADAPTATION TO NEW FOOD SOURCES IN TWO LABORATORY DOMESTICATED *CAENORHABDITIS ELEGANS* STRAINS

A Dissertation
Presented to
The Academic Faculty

by

Yuehui Zhao

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Biology in the
School of Biological Sciences

Georgia Institute of Technology
May, 2019

IDENTIFICATION OF THE GENETIC AND PHENOTYPIC BASIS OF ADAPTATION TO NEW FOOD SOURCES IN TWO LABORATORY DOMESTICATED *CAENORHABDITIS ELEGANS* STRAINS

Approved by:

Dr. Patrick McGrath, Advisor
School of Biological Sciences
Georgia Institute of Technology

Dr. Eric Gaucher
Department of Biology
Georgia State University

Dr. Joe Lachance
School of Biological Sciences
Georgia Institute of Technology

Dr. Levi Morran
Department of Biology
Emory University

Dr. Frank Rosenzweig
School of Biological Sciences
Georgia Institute of Technology

Date Approved: March 15, 2019

To my parents

ACKNOWLEDGEMENTS

First, I sincerely appreciate my doctoral thesis advisor Dr. Patrick McGrath who provides me solid and rigorous scientific training. His kind support and patience enable me to enjoy the fantastic evolutionary genetics research world. I also sincerely appreciate he trains me from both experimental and computational directions, which help me develop the necessary skills in quantitative bioscience and genomics. More importantly, I learned how to think critically and interpret the data thoroughly by Dr. McGrath.

I sincerely appreciate my doctoral thesis committee members Dr. Eric Gaucher, Dr. Joe Lachance, Dr. Levi Morran, and Dr. Frank Rosenzweig. All their kind guidance and help enable me to make the progress in my thesis research. Their valuable advice, encouragements, and supports help and inspire me a lot during my graduate study.

I also want to thank all the current and past members in McGrath's lab. We enjoy a great time in McGrath's lab. The collaboration with Wen Xu, Lijiang Long, Weipeng Zhuo, Dr. Richard Campbell, and Dr. Ed Large enable me to efficiently find the answer of the biological questions. With their kind help, I also learn a lot of essential experimental and computational skills. I also appreciate our lab manager Eric Anderson and technician Mark Lowder, they gave me the support for the experimental materials. I wish to say thank you to the past undergraduate students Akinade Ojemakinde, Caroline Page, Anchal Kamat. Their hard works help me to make the agar plates, measure the DNA concentration of the samples, and decontamination greatly supports my high-throughput competition experiment. I would like to thank Dr. Zack Johnson, Dr. Nicole Baran, Sanjeev Sariya and

all the other current and past lab members. Their critical review of my publications and the discussion with them teach me a lot.

I would like to appreciate my collaborators Dr. Fredrik Vannberg, Dr. Hang Lu, Jason Wan and Shweta Biliya at Georgia Tech; Dr. Erik Andersen, Dr. Steffen R Hahnel, Dr. Daehan Lee, Kathryn S Evans and Shannon C Brady at Northwestern University; Dr. Cori Bargmann and Dr. Joshua Greene at Rockefeller University. I learned a lot and broaden my horizon from these great collaborators.

I greatly appreciate the guidance, help, and great support from Dr. Francesca Storici and Dr. Annalise Paaby. I also thank the guidance and help from Dr. Yanyan Xing, Dr. Yuhong Fan, Dr. Chong Shin, Dr. Greg Gibson, Dr. Jeffrey Streelman, Dr. Kirill Lobachev, and Dr. Lin Jiang. I also appreciate other faculty members at Georgia Tech who gave me many bits of help during my graduate study.

I also appreciate all my dear friends at Georgia Tech and friends from other institutions. Their help and encouragement enabled me to make progress during my graduate study.

At last but not least, I sincerely appreciate all the support, encouragement from my dear parents and my dear family members. My mother Xiaomei Ju and my father Jian Zhao's unconditional love, encouragement, and support are my biggest motivation and the source of energy. They and other family members enable me to conquer all the challenges and complete this thesis research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ABZ	albendazole
ANOVA	analysis of Variance
ARL	allelic replacement line
BC	barcode
BSA	bulk segregant analysis
BZ	benzimidazoles
DAmP	decreased Abundance by mRNA Perturbation
E-MAP	epistasis mini array profile
EM	Expectation-maximization
GWAS	genome-wide association studies
GxE	gene-environment interaction
GxG	gene-gene interaction
HLE	Heated Liver Extract
HS-YE	HySoy-Yeast Extract
Indel	insertion/deletion
KO	knock-out
LD	linkage disequilibrium
LOD	log10 likelihood ratio
LTEE	long-term experimental evolution
MA	mutation accumulation
NIL	near-isogenic lines
QTL	quantitative traits loci

RIL	recombinant inbred lines
SGA	synthetic genetic array
SNP	Single-nucleotide polymorphism

SUMMARY

Genomic variations interact with the environment to drive the phenotypic diversity responsible for the vast biological diversity of the tree of life. Studying the genetic basis of the phenotypic variations and identifying the adaptive mutations can help us understand the mechanism responsible for trait differences including many human diseases such as cancer and pathogen infection. Characterizing the adaptive mutations and quantifying their fitness effects under different environment is non-trivial. Studying evolution in the lab using controlled experiments provides a route to deal with this challenge. In this doctoral thesis, I will first describe my study on the fitness effect of the derived "solitary" foraging behavior of N2 strain. N2 is a canonical strain widely used in more than 1,000 *C. elegans* labs across the world. This strain evolved the "solitary" foraging behavior caused by the two new mutations on the neural genes *npr-1* and *glb-5* Using competition experiments in response to specific environmental manipulations, I found although the derived alleles of *npr-1* and *glb-5* significantly increase animal's fitness, the derived solitary behavior is not responsible for the fitness benefits. Instead, I propose that the fitness benefits are caused by multiple biological traits including food consumption and reproduction. The behavioral change is a pleiotropic effect during evolution. Second, to identify additional adaptive mutations in N2 and LSJ2, I performed a high-throughput pairwise competition experiment using 89 N2 x LSJ2 recombinant inbred lines (RILs). I found an outlier RIL with higher fitness (RIL_{hf}) and more active exploration behavior than all the other RILs and two parental strains. To investigate if this is caused by higher-order epistasis or the *de novo* mutations, I crossed this RIL_{hf} to its two parental strains to create new RILs. Using these

new RILs, I performed a bulk segregant analysis with Illumina short-reads sequencing on behavioral phenotype and found a *de novo* complex genomic rearrangements at *rcan-1* gene, an ortholog of human Down Syndrome correlated gene RCAN1/DSCR1. The structure of this complex genomic rearrangement was resolved by using Nanopore long-read sequencing. This complex genomic rearrangement generates multiple version of *rcan-1* gene and is associated with fitness benefits and active exploration behavior. This provides us an experimental clue that how gene duplication occurs in metazoan and directly increases fitness benefits. In the last chapter, I will describe how the genetic variant interact with the environmental conditions regulate fitness. I performed pairwise competition experiments and quantitative traits loci (QTL) mapping using N2 x LSJ2 RILs in multiple laboratory environments including normal growth conditions, starvation conditions, and axenic liquid media. In addition, I also participated in research to study the fitness effects of genetic background under different temperature and study the fitness effects of drug resistance genetic variants under treatment with anthelmintic drugs. In sum, my doctoral research suggests that experimental evolution can serve as a powerful tool to study the genetic and phenotypic basis that regulate fitness in metazoan.

CHAPTER 1. INTRODUCTION

Most biological traits including many human diseases such as cancer, diabetes, neurological disorders, and aging have significant but complex genetic basis[1,2]. Identification of the causative genetic variants could help understand the biological mechanisms of diseases. The knowledge can be further used to predict and prevent the diseases as well as identify the targets for cures.

With the development of computational power and data science, geneticists are increasing their efforts to study the complex genetic basis of phenotypic variations. Understand the mechanism of how genetic factors and environmental factors regulate complex traits can give us more insights into the consequence of individual difference as well as contribute to the development of personalized medicine. However, studying the genetic basis of complex traits are non-trivial. The complex traits are regulated by a large number of genes that interaction with the environment. Understanding the mechanism requires to study gene-gene interaction (GxG) and gene-environment interaction (GxE)[3]. Despite the great contributions to study fundamental signaling pathways, the classical forward genetic approaches are biased to select single large effect mutation and the reverse genetic methods usually knock down/out or regulate up one gene or two genes to study phenotype under a controlled environment. It is difficult for the classical genetics tools to identify the variants with relatively small effects on the functional domain or *cis/trans* transcriptional elements, epistasis and gene-environment interactions that regulate complex traits. [4-7]

In a natural population, phenotypic variations are influenced by segregating genetic variants. The genetic variants arise, maintain, sweep or decay in the population are controlled by evolutionary forces including natural selection and genetic. Studying their roles in interacting with the environment, regulating phenotype, and affecting fitness are significant. Additionally, many genes have pleiotropic effects, it is not only important for us to identify the causal genetic variants, but also study which phenotype contributes to fitness effects. However, classical genetic tools lack the power to directly study complex interactions between genotype-environment-phenotype-fitness. Studying epistasis in natural populations remains challenging due to the statistical issues as well as many genetic variants segregate at low allele frequencies [8]. Furthermore, it is a general interest to identify adaptive mutations and understand how they regulate phenotypes to adapt to the local environment. In addition, identifying adaptive mutations and understanding the consequence of fitness increase can also give important insights into key questions in many diseases such as tumor clonal evolution and viral evolution. Classical genetics tools are unlikely to fully resolve these questions. Therefore, to bridge the gap of studying complex interaction of genotype-environment-phenotype-fitness, it is necessary to apply novel genetic tools in model organisms to bridge the gap of studying complex interaction of genotype-environment-phenotype-fitness (**Figure 1.1.1**).

In this introduction, I will first introduce the strategy and tools to study the genetic basis of phenotypic variations in this era of systems biology and next-generation sequencing. I will also give the examples of their usages to show the power in popular model organisms including *S. cerevisiae* and *C. elegans*. Then, I will introduce the recent researches that

applying experimental evolution and the derivative approaches to study evolutionary dynamics as well as some exciting results that elucidating the genetic basis of adaptation. At last but not least, I will introduce the two laboratory-adapted strains N2 and LSJ2. They are the major research objectives in my doctoral thesis. The main goal for this thesis studies adaptive mutations and understand how genetic variants and environmental factors to regulate *C. elegans* fitness. I believe the knowledge from this laboratory-adapted strains will give us many important insights into the principle of evolutionary genetics and disease process driven by evolution.

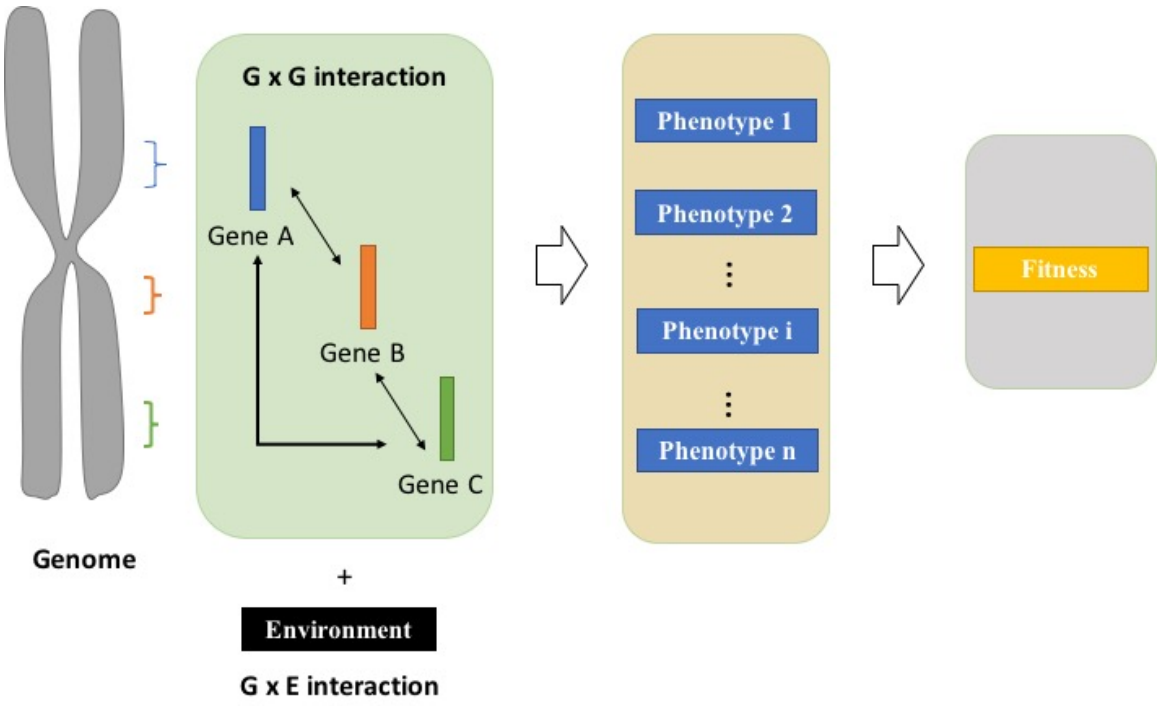

Figure 1.1.1. Interaction between genotype-environment-phenotype-fitness

1.1 Study genetic basis of phenotypic variations using systems biology tools

With the emergence and development of microarray, next-generation sequencing, and many high-throughput experimental platforms, genetic research is in an era of genomics systems biology. Statistics are also rapidly developed to fulfill the requirement to analyze and interpret the high-throughput data. Many powerful genetic approaches are developed and used to study genetic interactions and identify the genetic basis of phenotypic variations. I will briefly introduce some of the modern genetic tools and their usage examples in studying genetic interaction and identify the genetic basis of phenotypic variations in *S. cerevisiae* and *C. elegans.*

1.1.1 Synthetic Genetic Array and Epistasis miniarray profile

Using the advantage of budding yeast can exist as both haploid and diploid mode, researchers developed the powerful high-throughput approaches synthetic genetic array (SGA) and epistasis miniarray profile (E-MAP) to globally study yeast genetic interactions[9-11]. First, ~5000 library strains carry single nonessential gene deletion with mating type a is used to mate with the query strain carries a query mutation with mating type α. After sporulation and selection with selective markers, arrays of haploid strain carry double mutations are generated. To study global genetic interactions, ~5000 x ~5000 double mutant arrays can be generated by pin tool or Singer robot system. Then the arrays can be grown under a certain condition or with drugs and the colony sizes were quantified and further analyzed to study the genetic interactions of every two genes. These two methods are powerful to study globally genetic interactions. By comparing the colony sizes of parental strains carries single gene mutation and the colony sizes of double mutants carry

double mutations, the genetic interactions can be profiled and assigned into two groups: negative genetic interaction and positive genetic interaction. If we define the colony size of gene a deletion mutant *aΔ* is A, and the colony size of gene b deletion mutant *bΔ* is B. If the colony size of *aΔ bΔ* is smaller than A*B, the genetic interaction is assigned to negative interaction. The negative genetic interaction refers to the double mutations have a synthetic effect on fitness that implies they may play roles in redundant pathways. The most representative example of a negative effect is double mutations causes synthetic lethality. If the colony size of *aΔ bΔ* equals A or B, the genetic interaction is assigned as a positive interaction. The positive genetic interaction implies the two genes play roles in the same pathway and may build the same protein complex. The derivative experimental strategy is developed in other systems such as mammalian cell culture[12].

Despite powerful to systematically study the genetic interactions, SGA and EMAP are mostly using gene deletion mutant library, Decreased Abundance by mRNA Perturbation (DAmP) mutant library[13]. The genetic interactions studied by SGA and EMAP are limited in loss of function mutations that not sufficient to represent the various types of natural variants. For instance, it is difficult for the two approaches to study gain of function mutations that not delete the whole ORF or 3'UTR or the mutations located on regulatory elements.

1.1.2 Quantitative traits loci QTL using linkage mapping

To understand the genetic basis of phenotypic variation, quantitative traits loci (QTL) mapping is a powerful tool. It detects the genome region with genetic markers (eg. single polymorphism marker) that correlate with the phenotypic difference. There two major

methods to mapping QTLs, the one is linkage mapping which 1) maps QTL in families from the outbred population or 2) maps QTL in segregating offspring from inbred lines. The other is association mapping that maps QTLs in random mating populations. Here I will mainly describe the linkage mapping using segregating generation from inbred lines. This is a major method used in model organism [14,15].

Applying linkage mapping first require construction of a large number of inbred lines by crossing two parental strains with a desired phenotypic difference. F1 heterozygotes could be backcrossed with one of two parental strains or intercrossed with each other to generate F2 individuals. By inbreeding F2 for 10-20 generations, the recombinant inbred lines (RILs) are constructed and ready for phenotype quantification. The genotyping information of each RIL can be characterized by next-generation sequencing or other high-throughput genomics approaches. The desired phenotypes can be quantified of the RILs. Once we obtain the genotyping information and phenotyping information of all the RILs, the single marker analysis or interval mapping can be performed to calculate the LOD score ($\log_{10}$ likelihood ratio). Then the significant threshold of LOD can be determined by the permutation test. Single marker analysis is doing Analysis of Variance (ANOVA) and calculating the F statistic to determine the LOD score. This method is simple but excludes the sample with missing genotype. The interval mapping method is popular since it conquers this question using the EM algorithm (Expectation-maximization)[14,15].

QTL is successfully used to resolve many complex traits in multiple model organisms. For example, QTL mapping is used to study DNA repair sensitivity and the genetic factors regulate genome instability[16,17]. In *C. elegans*, QTL mapping is successfully performed to

identify the genetic basis that regulates phenotypic difference aerotaxis, foraging behavior, and avermectin resistance, [18-20].

1.1.3 Genome-wide association (GWA) mapping

Unlike linkage mapping, Genome-wide association mapping is using a randomly mating population. Utilizing the historical recombination in the population, the genetic loci correlates with a phenotypic difference could be identified. This approach is based on linkage disequilibrium (LD) between the genetic loci correlated with phenotype and the genetic markers[14]. The chi-squared test could be performed to comparing the difference of expected individual number with the genotype and observed individual number with the genotype. If the difference is statistically significant, it suggests the loci affecting traits are in LD with marker locus.

This approach is widely used in modern human genetics to identify loci associated with diseases and other traits[21,22]. In the model organism, this method is also widely used in multiple model organisms. For instance, it is successfully applied in *Drosophila* species to study cold tolerance and body size in seasonally varying environment. It also used to measure *Arabidopsis thaliana* fitness traits under different environment.[23,24]. In the *C. elegans* research community, Dr. Erik Andersen and his colleagues collect *C. elegans* strains from the worldwide and genotype all these wild isolates by next-generation sequencing. With these resources, it provides valuable resources to study the genotype and phenotype correlation using GWA mapping [25]. One example is using GWA to study *C. elegans* fitness in different climate environment [26].

1.1.4 Bulk Segregant Analysis (BSA)

With the development of next-generation sequencing, bulk segregant analysis (BSA) is applied to identify the causal genetic variants in two strains, which lead to a phenotypic difference. After crossing the parental strains with a desired phenotypic difference, a large amount of progeny (F2 or later generation) are used for recording phenotypes. The progeny with low phenotypic value are pooled together and the progeny with high phenotypic value are pooled together. The genomic DNA is isolated from low phenotype group and high phenotype group for whole genome resequencing. The allele frequency is quantified in each group. The allele frequency close to 0.5 suggests the genetic loci are not responsible for the phenotypic difference, while the genetic loci show the difference of allele frequency suggests the correlation of genotype and phenotype. In general, more generations enable more recombination events to happen during meiosis and a larger number of progeny can result in a higher resolution to characterize the causal genetic loci. The experimental design should consider both resolution and time/cost. BSA analysis shows strong power to identify multiple QTL in budding yeast, fruit fly, plants[27-31]. This approach is also developed and utilized in *C. elegans*. *fog-2* mutant affects germ-line sex determination and transforms hermaphrodites to males. Using *fog-2* mutant allows researchers to generate a large number of segregants. One example is using this approach to study drug resistance[32].

1.2 Study genome dynamics and adaptive mutations using experimental evolution

Using the modern quantitative genetic and genomic methods described above enable us efficiently to study complex traits. However, this is still not sufficient to fully draw the blue-script of the evolution of these phenotypic differences. What the evolutionary history

of the genome and phenotypes? What's the potential evolution route of the genome and phenotypes? What's the rule of evolutionary dynamics? Studying adaptive mutation is one of the key questions in evolutionary genetics. Studying the fitness effect of the genotype changes and the phenotype changes is crucial to understand how the organisms adapt to the environment. Experimental evolution serves as a valuable method to answer these questions.

In brief, experimental evolution is to study a population across multiple generations in a defined experimental condition. The research includes but not limited in 1) the changing rates of genotype and phenotype, 2) the divergence of genotype and phenotype of the population, 3) the interaction between genotype-phenotype-environment to regulate fitness, 4) the repeatability of evolution [33]. Utilizing experimental evolution to test hypothesis can be traced back to the late 19th century when William Dallinger performed his uncompleted evolution experiment from 1880 to 1886. The flagellates were cultured in an incubator with slowly increasing temperature from 60 Fahrenheit to 168 Fahrenheit. The flagellates are adapted to the higher temperature but cannot survive in the optimal temperature[34]. With the development of modern molecular biology and next-generation sequencing, experimental evolution shows great power to study the evolutionary process in the unicellular organism and metazoan. I will briefly introduce some representative experimental evolution researches.

1.2.1 Experimental evolution in unicellular model organisms

One of the most representative experimental evolution research is launched by Dr. Richard Lenski, the long-term experimental evolution with *Escherichia coli* (LTEE). The project

started on February 24, 1988, and still ongoing. 12 populations of two strains are cultured in a glucose-limited minimal salts medium. The medium supports 5×10^7 cells/ml, and samples are stored at -80°C every 500 generations. Approximately 6.67 generations pass through every day, making more than 70,000 generations[35]. This experiment enables researchers to directly monitor the dynamics of the evolutionary process. Using next-generation sequencing and other genotyping technology, the researchers could monitor the genome dynamics across the evolution[36] (**Figure 1.2.1**). By performing a competition experiment with ancestor strain, the relative fitness of the evolved strain can be directly measured. In addition, many intriguing evolutionary phenomena can be monitored that include but not limit the evolution of the new trait, hypermutator, clonal interference, parallel evolution, and second-order selection[37-40].

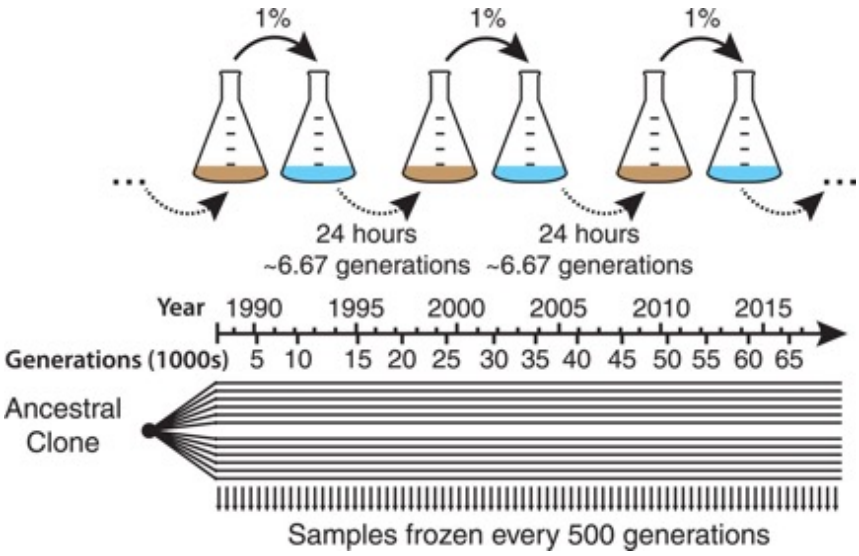

Figure 1.2.1. Scheme of LTEE. Figure source: Blount, Z. D., Lenski, R. E. & Losos, J. B. Contingency and determinism in evolution: Replaying life's tape. *Science* **362**, eaam5979 (2018).[35]

Another popular unicellular model organism utilized in experimental evolution is budding yeast. The researchers using automating platform perform high-throughput evolutionary experiment found pervasive hitchhiking and clonal interference phenomenon[41]. Using yeast also enable scientists to study epistatic effects regulate fitness [42]. In recent years, the researchers developed the barcoding lineage tracing system, which can be used to trace the fate of the lineage in high-throughput scale[43,44]. This system does not only show the great power to study the adaptation process, but also show the great potential to characterize the beneficial mutations. This system is also applied to *E. coli* for studying the evolutionary dynamics[45].

There are many other intriguing experimental evolution projects are launched. The researchers have inserted the engineered ancestry elongation factor into *E. coli* and applying the experimental evolution to gain insights into historical adaptive process[46]. By reconstructing the ancestral sequence, the researchers built an experimental phylogeny for studying the mechanisms of molecular evolution[47]. Using yeast and algae, the researchers are investigating how unicellular organisms evolve and transform into multi-cellular organisms [48,49]. The experimental evolution shows great power to understand the evolution of phenotypic difference, study complex interaction between genotype-phenotype-environment-fitness, as well as the theoretical population genetics. The evolutionary experiment takes a long time and costs to perform. Thus, the experiments should be designed very carefully[33].

1.2.2 Experimental evolution in metazoan

Experimental evolution is not only launched in a unicellular organism but also in the metazoan system. One popular model organism in experimental evolution are nematodes *C. elegans* and *C. remanei.* Nematode system has many advantages in biological research: 1. well sequenced and annotated genome; 2. short life cycle with self-fertilization hermaphrodite lifestyle; 3. many biological pathways are more conserved as human's; 4. efficient molecular biological techniques; 5. easy to feed and culture as large population size. With these benefits, it serves as a metazoan experimental evolution model to study population genetic theories and specific evolutionary biological questions. Mutation accumulation experiments (MA) were performed to characterize the mutation rates and mutation types occurred in *C. elegans*. One study of MA lines adaptive recover experiments using next-generation sequencing detected selective sweeps with parallel mutations[50]. Another study using MA lines found a rapid increase of frequency in gene copy-number variants[51]. Work using a strain with beneficial allele invaded population with deleterious allele experimentally tested Haldane's prediction that extinction of beneficial allele by genetic drift declines with a number of invading beneficial alleles[52]. Experimental evolution in *C. elegans* also implies the opportunity to study balancing selection[53]. Recent work of outcrossing with wild isolates and experimental evolving the generated RILs identified a couple of strong epistatic interactions. Besides, experimental evolution also applied to address specific questions including development, stress resistance, aging, reproducibility, and sex in *C. elegans* [54-57]. Similar evolution experiments also carried out by using another nematode species *C. remanei* [58,59]. Additionally, *C. elegans* also serve as a good model to study post-pathogen interaction. Work of coevolution with pathogen bacteria *Serratia marcescens* tested and experimentally supported Red-Queen hypothesis that the persistence of out-crossing facilitated by selection from coevolution of pathogens [60].

The other popular model organism is *Drosophila melanogaster*. One of the most representative work is a *D. melanogaster* continually under laboratory selection for accelerated development or body size over several hundred generations. These studies give us more insights into the evolution of development and aging[61-63]. Using metazoan to carry out experimental evolution can help us to learn more about the evolution of the complex biological processes such as development and behavior.

1.2.3 Derivative approach to study evolutionary dynamics in tumor clonal evolution and viral evolution

Evolutionary experiments are also utilized in tumor clonal evolution and viral evolution. The researchers utilize single-cell DNA sequencing approach to monitor the genome dynamics of copy number changes of tumor clone [64,65]. This powerful technology enables the researchers to monitor evolutionary history during the cancer processes such as early stage of breast cancer and chemoresistance [66,67]. The researchers found the mixture of tumor clones rather than single tumor clone becoming invasive in early-stage of breast cancer. Another interesting finding is that chemoresistance mutation have already existed a long time before chemotherapy. The similar idea using the evolutionary experiment to study the cancer process is also used in other cancer types such as bladder cancer and lung cancer[68]. Studying cancer process from an evolutionary perspective can give us more insights into the mechanism of cancer initiation, metastasis, and resistance for cancer therapy.

The researchers also used the same idea to detect the evolutionary dynamics of influenza viral infection in four cancer patients. They detected clonal interference during infection

of the patients[69]. The knowledge from this study can lead to researchers better understand the evolution process of influenza, the potential evolving infection strategy, and the resistance to vaccines. Overall, these biomedical researches can expand human's knowledge on the mechanisms of many diseases like cancer, AIDS and flu. This type of evolutionary research can help us identify effective therapeutic approaches.

1.3 Two laboratory-adapted *C. elegans* strains N2 and LSJ2

The laboratory-adapted *C. elegans* strains N2 and LSJ2 were domesticated in laboratory condition. Using these two strains help *C. elegans* biologists elucidate the mechanisms the complex traits including aerotaxis and development. I will first describe the domestication history of the two strains. Then, I will introduce the work that study fixed mutations in the two strains lead to insights into the mechanisms of complex traits.

1.3.1 Laboratory domestication of N2 and LSJ2 – an unintentional long-term experimental evolution in C. elegans

N2 and LSJ2 share a common ancestor following its isolation from the wild in UK Bristol 1951 (**Figure 1.3.1a**)[70,71]. The ancestor strain was collected from mushroom compost and then the strain was isolated by Warwick Nicholas. The Bristol *C. elegans* strain was then transported to the laboratory of Ellsworth Dougherty at the Kaiser Foundation Research Institute in Richmond, CA. Between 1957 and 1958, the Bristol ancestor strain was cultured in two types of growth condition: (1) agar plates seeded with bacteria as a food source; (2) heated liver extract (HLE) axenic liquid media.

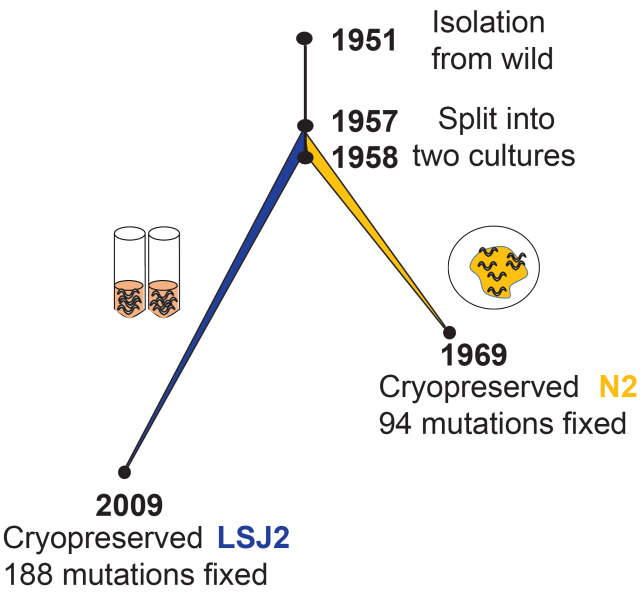

Figure 1.3.1. Laboratory domestication of N2 and LSJ2. The ancestor of the two strains is isolated in 1951 at Bristol, UK. The strain was split into two different cultures (1) agar plates seeded with bacteria as a food source; (2) Liquid axenic media with heated liver extract media as a food source. The lineage cultivated in agar plate was frozen in 1969 and named as N2. The lineage cultured in liquid media was frozen in 2009 and named as LSJ2.

In the 1960s, Sydney Brenner tried to find an organism that can be used to study more complex biological questions such as development and behavior. In a letter, he wrote: "I would like to tame a small metazoan organism to study development directly". [72] Since the nematode could be easily grown in the lab and a small number of neurons, he considered nematode as the target for his developmental and behavioral study. In 1963, he requested the Bristol ancestor strain from Dougherty's lab. The strain was sent to the laboratory of Molecular Biology at Cambridge and Sydney cultured the animals on agar plates seeded with *E. coli*. The strain was cryopreserved in 1969 and named as N2. N2 was then used for

subsequent forward genetics study on behavior in Brenner's Laboratory. In 1974, Sydney introduced N2 as a model to study genetics to biology community [73]. Today, this strain is widely used in more than 1000 labs across the world and become the canonical strain in *C. elegans* research labs[74]. The other lineage in axenic HLE liquid media was used for studying nutrient metabolism. This lineage was firstly frozen in 1995 and named as LSJ1. This lineage was then continually cultured in HLE liquid media for 14 years and then was frozen in 2009 and named as LSJ2 [71].

During its laboratory domestication, N2 spent 300-2000 generations between 1951-1969 [70]. LSJ2 was transferred every month between 1958-2009. Thus, LSJ2 has transferred at least 600 generations, although more than one generation likely occurred in a single month. During the laboratory growth of the two strains, the researchers noticed that evolution is going on and the two strains show many phenotypic variations. Using whole genome resequencing, McGrath identified 94 new mutations are fixed in the N2 lineage and 188 new mutations fixed in the LSJ2 lineage [71]. Similar to Richard Lenski's LTEE, the laboratory domestication of N2 and LSJ2 serves as an unintentional long-term metazoan experimental evolution model. Unlike LTEE or experimental evolution using budding yeast or green algae, N2 and LSJ2 could be used to test adaptive hypotheses in a multicellular organism. The two strains provide valuable resources to study the complex interaction of genotype-environment-phenotype-fitness.

1.3.2 Identified fixed mutations regulate phenotypic variations

Some fixed mutations in N2 and LSJ2 were characterized by different labs. I will introduce the fixed mutations in N2: *npr*-1, *glb*-5, and *nath-10* as well as the fixed mutations in LSJ2: *srg-36, srg-37* and *nurf-1*.

The most representative example is the N2 derived alleles of neuronal genes *npr-1* and *glb-5*, which lead to animals foraging behavior changes. The animals carry ancestral alleles of *npr-1* and *glb-5* display aggregation on the border of bacteria lawn in the agar plates. This type of foraging behavior also called social behavior. On the other hand, the animals that carry N2 derived alleles of *npr-1* and *glb-5* do not aggregate together and display solitary behavior. *npr-1* encodes a NPY family neuropeptide receptor gene and the derived allele of *npr-1* has a point mutation leading to 215th Phe changes to Val. The role of *npr-1* in regulating social behaviors was firstly characterized by Mario de Bono in Cori Bargmann's lab at Rockefeller University [75]. *glb-5* is a neural globin gene and the derived allele of *glb-5* has a 765bp duplication structural variant which induces a premature stop codon [19]. *glb-5* was firstly characterized by the quantitative analysis of the behavioral response to O_2 level changes [19,76]. These two alleles and the behavioral difference serves as a model to study social behavior. After characterizing *npr-1,* many projects are launched in the de Bono's lab, Bargmann's lab and other labs have revealed the neural circuit, molecular mechanism, and genetic basis of this gene. Researchers found natural O_2 level of *C. elegans* is around 10% and the aggregation and bordering social behavior mainly for wild animals is to avoid high ambient 21% O_2 (**Figure 1.3.2**). In addition to cGMP-gated channel *tax2/tax-4* and soluble guanylate cyclase *gcy-35* and *gcy-36*, *glb-5* also plays a role in facilitating O_2 sensing in URX, AQR, and PQR oxygen sensing neurons [5,19,76,77]. When the O_2 level is high, URX, AQR, and PQR are activated which stimulate the activation of

inter/motor neuron RMG and lead to animals aerotaxis and social foraging behavior to avoid high level O_2 [78]. In addition, *glb-5* expressed in BAG neurons, which sense the CO_2 level changes. *npr-1* senses the neuropeptide FLP-21 and found that plays a role suppressing RMG activity. The derived allele of N2 *npr-1* 215V is a gain-of-function mutation that lead to sensitivity to FLP-18 and suppresses the social behavior under ambient O_2 level[79]. In addition to regulating O_2 sensing and pheromone sensing, URX also plays role in sensing nutrients in the intestine to regulate metabolism[80]. The more details of URX-RMG neuronal circuit to regulate social foraging behavior will be discussed in Chapter 2.

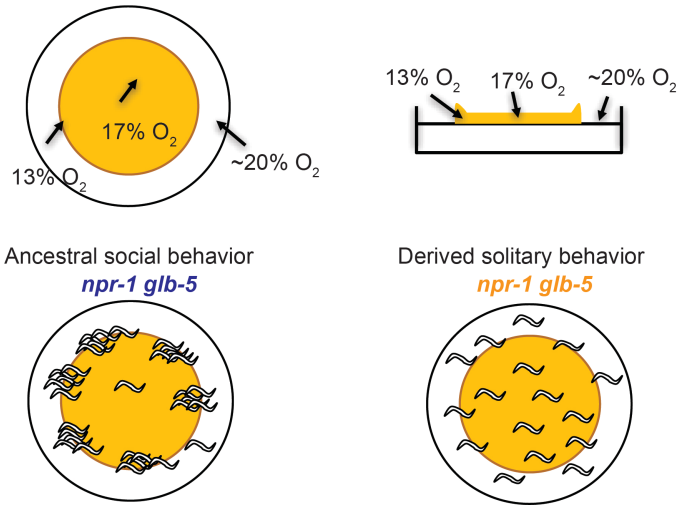

Figure 1.3.2. Ancestral social behavior and derived solitary behavior. Blue color: ancestral alleles of *npr-1* and *glb-5*; Orange color: derived alleles of *npr-1* and *glb-5*.

Another N2 derived allele *nath-10* M746I, Human N-acetyltransferase homolog gene *NAT10*, was found to affect vulval cell-fate specification[81]. This derived mutation located

in the conserved N-acetyltransferase domain. The derived allele also shows affects fecundity and growth rate. This derived allele is reported to show fitness benefits in a competition experiment.

LSJ2 has fixed more mutations[71]. Comparing to N2 lineage, there are many phenotypic differences in LSJ2 lineages. Under high pheromone concentration with a limited food source, N2 progeny enter into dauer stage while LSJ2 progeny do not. This dauer pheromone tolerance is due to a large genomic deletion at Chromosome X which disrupt two G-protein coupled receptor genes *srg-36* and *srg-37*. These two genes are primarily expressed in ASI chemosensory neurons to regulate dauer pheromones[71]. The LSJ2 lineage also shows delayed egg laying timing. Using a QTL mapping of egg-laying timing experiment, a strong significant QTL was found at Chromosome II right arm including a 60bp deletion at C-terminus of chromatin remodeling factor *nurf-1*, an ortholog of human chromatin remodeling factor BPTF. Using CRISPR/Cas9 genome editing, the derived allele of *nurf-1* 60bp deletion was found to regulate egg laying rates[82,83].

There are many phenotypic changes in these two laboratory domesticated strains. These phenotypic changes may or may not interact with the laboratory growth environment to contribute to fitness benefits. It will be intriguing to identify the beneficial mutations and test the adaptive hypothesis.

1.4 Research aims

In this doctoral thesis, I have three major research aims as below:

Specific Aim1: Study if the derived solitary foraging behavior in N2 lineage has fitness benefit in standard laboratory growth condition.

Derived *npr-1* and *glb-5* in N2 lineage lead to ancestral social foraging behavior transforms to solitary foraging behavior. In this aim, I firstly verified if the derived alleles of *npr-1* and *glb-5* are fixed due to natural selection or genetic drift. Then, by modifying the oxygen levels to suppress ancestral social behavior, I found the behavior difference does not necessary for the fitness advantage. Using RNA-seq and multiple physiological assays, I found there are other biological traits also changed by the two derived alleles. Oxygen sensing neuron URX plays a major role in fitness effects of the derived alleles of *npr-1* and *glb-5*.

Specific Aim2: Study which new mutations in N2 and LSJ2 are adaptive in standard laboratory growth condition.

There are around 300 mutations are fixed in N2 and LSJ2 lineages. To characterize if there are other mutations increase fitness in standard laboratory growth condition in addition to *npr-1* and *glb-5*, I carried out a high-throughput fitness assay using N2/LSJ2 recombinant inbred lines (RILs). In this assay, one RIL shows higher fitness than parental strains and all other RILs. This RIL also displays active exploration behavior. To investigate if the fitness increase and behavioral change are caused by higher-order epistasis or *de novo* mutation, I launched bulk segregant analysis and mapped a *de novo* complex genomic rearrangement at *rcan-1* gene region. Using long-read Nanopore sequencing, the complex genomic rearrangement structure is resolved. The result of the competition experiment and behavioral assay suggest it is the casual mutation lead to fitness and behavioral change.

Specific Aim3: Using competition experiment to study fitness effect under different environmental conditions

It is of a general interest to understand how environmental factors interact with genetic components to regulate biological traits. In this aim, I used competition assay under different environmental factors to study how gene-environment interaction (GxE) regulate *C. elegans* fitness. These experiments including using N2/LSJ2 RILs to perform competition assay in 1) starving growth condition in agar plates; 2) Heated liver extract media; 3) Different temperature environment; 4) agar plates plus anthelmintic drug benzimidazoles.

CHAPTER 2. CHANGES TO SOCIAL FEEDING BEHAVIORS ARE NOT SUFFICIENT FOR FITNESS GAINS OF THE *C. ELEGANS* N2 REFERENCE STRAIN

Changes to social feeding behaviors are not sufficient for fitness gains of the *C. elegans* N2 reference strain

Yuehui Zhao[1], Lijiang Long[1], Wen Xu[1], Richard F. Campbell[1], Edward E. Large[1], Joshua S. Greene[2], and Patrick T. McGrath[1,3,4]

1. Department of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA
2. The Rockefeller University, New York, NY 10065, USA
3. Department of Physics, Georgia Institute of Technology, Atlanta, GA 30332, USA
4. Institute of Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, GA 30332, USA

2.1 Abstract

The standard reference *Caenorhabditis elegans* strain, N2, has evolved marked behavioral changes in social feeding behavior since its isolation from the wild. We show that the causal, laboratory-derived mutations in two genes, *npr-1* and *glb-5*, confer large fitness advantages in standard laboratory conditions. Using environmental manipulations that suppress social/solitary behavior differences, we show the fitness advantages of the derived alleles remained unchanged, suggesting selection on these alleles acted through pleiotropic traits. Transcriptomics, developmental timing, and food consumption assays showed that N2 animals mature faster, produce more sperm, and consume more food than a strain containing ancestral alleles of these genes regardless of behavioral strategies. Our data suggest that the pleiotropic effects of *glb-5* and *npr-1* are a consequence of changes to O2 -sensing neurons that regulate both aerotaxis and energy homeostasis. Our results demonstrate how pleiotropy can lead to profound behavioral changes in a popular laboratory model.

2.2 Introduction

It is tempting to compare the endless forms of life and create adaptive hypotheses to explain their differences. Why are polar bears white? Perhaps as camouflage for when they hunt. Or perhaps to make it easier to absorb heat from the sun. Both explanations make sense, but designing experiments to distinguish between these possibilities is not trivial. Further, as Gould and Lewontin critiqued, relying on adaptive evolution as the sole explanation for phenotypic change while ignoring alternative explanations such as genetic drift, adaptive

constraints, or pleiotropy does not follow Darwin's pluralistic approach [84]. In the current era, inexpensive next-generation sequencing and increasingly sophisticated bioinformatics analysis enable the identification of causative mutations with signatures of selection, yet it is difficult to determine *why* these alleles are under selection. Indeed, the pervasive effects of pleiotropy means that signatures of selection alone are not enough, adaptive hypotheses must be tested directly. Experimental evolution offers a route to test hypothesis directly [85-87]. The ability to manipulate model organisms in the lab provides a greater opportunity to test adaptive hypotheses beyond arguments of plausibility and address the role of these other competing themes in the evolution of biological traits.

These studies are also useful for understanding how organisms adapt to laboratory conditions. Since the fundamental work of Gregor Mendel elucidating the laws of genetic transmission, model organisms have enabled experimenters to gain fundamental insights into many biological processes. Modern research tools are facilitating the use of new and unusual species to analyze longstanding biological questions [88-91]. More and more species are reared in the laboratory as models for biological traits of interest. An issue for these approaches, particularly for comparative analysis or for those addressing evolutionary questions, is the extreme shift in environment and associated selective pressures that these populations experience. All species evolve through the process of natural selection and genetic drift; many model organisms have evolved by exposure to the novel and artificial conditions experienced in the lab [81,92-97]. Understanding the process of adaptation of wild populations to captivity is necessary to understand how the genetic, developmental, and neural circuits are changed in these laboratory populations.

As a model for understanding laboratory adaptation in a multicellular organism, we have focused our studies on the N2 strain of *Caenorhabditis elegans*. N2 is the canonical reference strain used by hundreds of *C. elegans* labs across the world. While this strain was introduced to the genetics research community by Sydney Brenner in 1974 [73], it was actually isolated by L.N. Staniland and Warwick Nicholas from mushroom compost in 1951, spending multiple decades (~300-2000 generations) in two primary growth conditions: on agar plates where bacteria was its primary food source or in liquid axenic media [70]. A small number (~100) of new mutations that arose and fixed in the N2 strain following isolation from the wild have been identified [71], including a neomorphic, missense mutation in the neuropeptide receptor gene *npr-1* and a recessive, 765 bp duplication in the nematode-specific globin gene *glb-5*. These mutations were originally identified for their role in foraging and aerotaxis behaviors and were initially thought to represent natural genetic variants [75,76] (**Figure 2.1.1**). A large body of work has found that these genes regulate the activity of the URX-RMG neuronal circuit that controls O_2 responses on food [5,19,76-78,98]. Animals with the ancestral alleles of *npr-1* and *glb-5* prefer ~10% O_2 concentrations while foraging and follow O_2 gradients to the border of bacterial lawns (~12% O_2) and feed in groups (called social behavior); animals containing the derived alleles of these genes are less sensitive to 10-21% O_2 gradients in the presence of food and feed alone (called solitary behavior) [5,77,99].

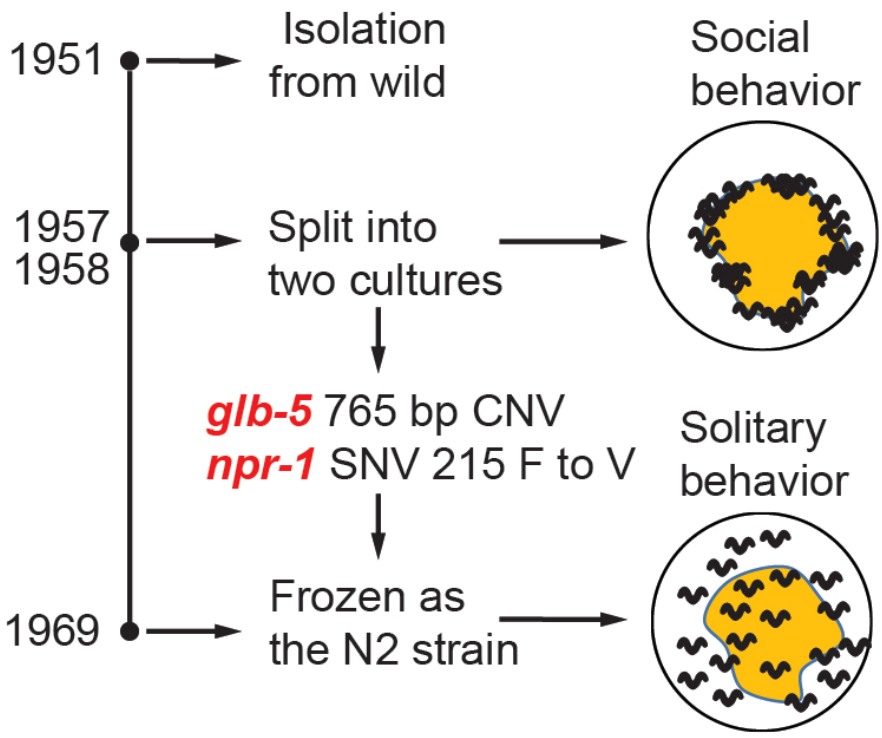

Figure 2.2.1 Overview of life history of the standard reference N2 strain since its isolation from the wild. Derived alleles in npr-1 and glb-5 arose and fixed after 1957 and before 1969 when methods for cryopreservation were developed. These two alleles were identified for their role in changing foraging behavior on bacterial lawns from social to solitary behavior.

We have previously proposed that the derived alleles of *glb-5* and *npr-1* were fixed by selection as solitary animals are more likely to be picked when propagating animals to new plates [19]. However, aggregation behavior in the ancestral *npr-1* strain appears to create local food depletion leading to a weak starvation state, which reduces reproduction and growth [100]. Potentially this starvation difference could be responsible for the fitness differences of the strains. Consistent with both hypotheses, a number of experimental

crosses or competition experiments between parental strains that are polymorphic for *npr-1* have resulted in enrichment of the derived allele of *npr-1*, suggesting it confers a fitness advantage under standard lab husbandry [101-103].

In order to distinguish between these hypotheses, we performed pairwise competition experiments following a number of environmental and/or genetic manipulations. Surprisingly, our results suggest that neither hypothesis is correct. While the derived *npr-1* and *glb-5* alleles increase fitness of animals on agar plates, the differences in social vs. solitary behavior are not necessary for their differences in fitness. Instead, our work suggests that fitness gains are due to increases in food consumption and changes in reproductive timing, mediated by O_2-sensing body cavity neurons that are also required for social feeding behaviors. Our work demonstrates that even when alleles are identified that confer fitness advantages, care must be taken in inferring the phenotypes that are responsible due to the pleiotropic actions of genetic changes.

2.3 Results

2.3.1 Derived alleles of npr-1 and glb-5 increase fitness in laboratory conditions

In previous reports, we have used multigenerational pairwise competition experiments to compare the relative fitness of two strains (**Figure 2.3.1a**) utilizing Droplet Digital PCR with a custom TaqMan probe to quantify the proportion of each genotype [18,104,105]. To quantify this change, we used a generic selection model to estimate the relative fitness difference (w) between the two strains (**Figure 2.3.1b**). In this context, relative fitness

measures the generational change in relative abundance of each of the two strains. We also used CRISPR-enabled genome engineering to create strains with a silent mutation in the *dpy-10* gene using a previously published high-efficiency guide RNA (**Figure 2.3.1c**) [106], which we will refer to as barcoded strains. These strains allow us to use a common Taqman probe to quantify the relative fitness of a test strain against these barcoded strains. We confirmed that the *dpy-10* silent mutation had no statistically significant effect on fitness in two genetic backgrounds studied throughout this report (**Figure 2.3.1d**).

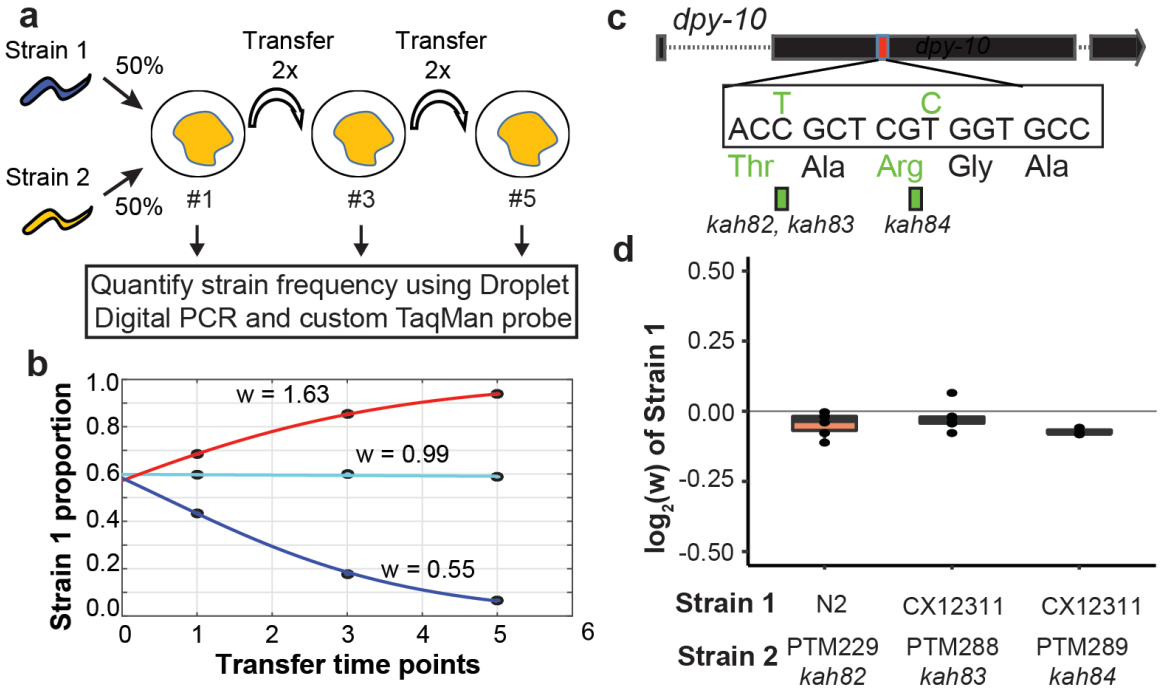

Figure 2.3.1 Schematic of competition assays used to measure relative fitness levels between two strains. **a**. Schematic of pairwise competition experiments used throughout the paper to quantify fitness differences between two strains. **b**. Relative proportion of each strain as ascertained by Droplet Digital PCR using a custom TaqMan probe (dots) is used to estimate the relative fitness between the two strains (line). **c**. Silent mutations were edited into the 90th or 92nd amino acid of the *dpy-10* gene using CRISPR/Cas9 to create a

common SNV for Droplet Digital PCR. We refer to these as barcoded strains. **d**. Competition experiments between the parent strain (top) and the same strain containing one of the silent mutations. We display the result from each competition experiment as a single dot overlaid on top of a boxplot showing the mean, first, and third quartiles of all replicates.

In order to test the fitness effect of the derived alleles of *npr-1* and *glb-5*, we utilized three previously described near isogenic lines (NILs) containing ancestral alleles of *npr-1* (QG1), *glb-5* (CX10774), or both genes (CX12311) introgressed from the Hawaiian CB4856 wild strain into the standard N2 background [19,71,107]. The *npr-1* introgressed region is ~110 kb in size and the *glb-5* introgressed region is ~290 kb in size. For brevity, we will refer to genotype of these introgressed regions throughout this report by the ancestral/derived allele they contain (e.g. the ancestral allele of *npr-1* vs the introgressed region containing the ancestral allele of *npr-1*). For clarity, we will refer to the NILs colloquially using the ancestral introgression(s) they contain instead of their opaque strain names (i.e. N2 = N2, CX10774 = $N2_{glb\text{-}5}$, QG1 = $N2_{npr\text{-}1}$, and CX12311 = $N2_{glb\text{-}5,\ npr\text{-}1}$). If needed, readers can find the strain name used in each figure in the supplemental source data files. In contrast to the N2 strain, the $N2_{glb\text{-}5,\ npr\text{-}1}$ strain aggregates at the border of bacterial lawns where O_2 levels are lowest due to the increased height of the bacterial lawn. We confirmed previous reports that both the derived alleles of *npr-1* and *glb-5* suppress bordering behavior to varying degrees [19,75,108]; *npr-1* accounted for the majority of the difference with *glb-5* playing a modulatory role (**Figure 2.3.2a**). To compare the relative fitness of the four strains, we competed each strain against the barcoded $N2_{glb\text{-}5,\ npr\text{-}1}$ strain,

transferring animals each generation by washing to minimize potential sources of investigator bias towards picking social or solitary animals (**Figure 2.3.2b**). The N2 strain was the most fit in these conditions, with a relative fitness (w) of ~1.30. Interestingly, the fitness effects of the *glb-5* and *npr-1* regions were epistatic - the derived allele of *glb-5* increased the relative fitness in the derived *npr-1* background but showed no effect in the ancestral allele of *npr-1*. The derived *npr-1* allele increased fitness in both backgrounds of *glb-5*. To confirm the fitness advantage of the derived *glb-5* allele in the derived *npr-1* background, we also competed the $N2_{glb\text{-}5}$ strain against the barcoded N2 strain (**Figure 2.3.2b**). The estimated selective coefficient (a common measure of the fitness difference of a beneficial allele) of the *glb-5* allele in the *npr-1* derived background was s = 0.10 (0.06 – 0.13 95% confidence interval), the estimated selective coefficient of the *npr-1* allele in the *glb-5* ancestral background was s = 0.17 (0.12 - 0.23 95% confidence interval), and the estimated selective coefficient of the *npr-1* allele in the *glb-5* derived background was s = 0.30 (0.27 - 0.34 95% confidence interval). These selective coefficients are comparable to beneficial alleles identified in other organisms, such as the haplotype responsible for lactase persistence (~0.01-0.19) [109] and the sickle-cell trait (0.05 – 0.18) in humans [110].

While the introgressions surrounding the *npr-1* and *glb-5* genes are relatively small, these NIL strains carry additional polymorphisms in surrounding genes from the CB4856 strain. We also performed competition experiments using two previously published *npr-1* loss-of-function alleles *(ad609* and *ky13)* [75] against the N2 barcoded strains. Both the *npr-1(ad609)* and *npr-1(ky13)* loss-of-function alleles decreased the animal's relative fitness in an amount comparable to the ancestral allele (**Figure 2.3.2c**). We did not perform similar experiments on the *glb-5* gene. Altogether, our work suggests that the *npr-1* derived allele

increases fitness of animals in laboratory conditions and also suggests that the derived allele of *glb-5* increases the fitness of animals in a *npr-1* dependent manner.

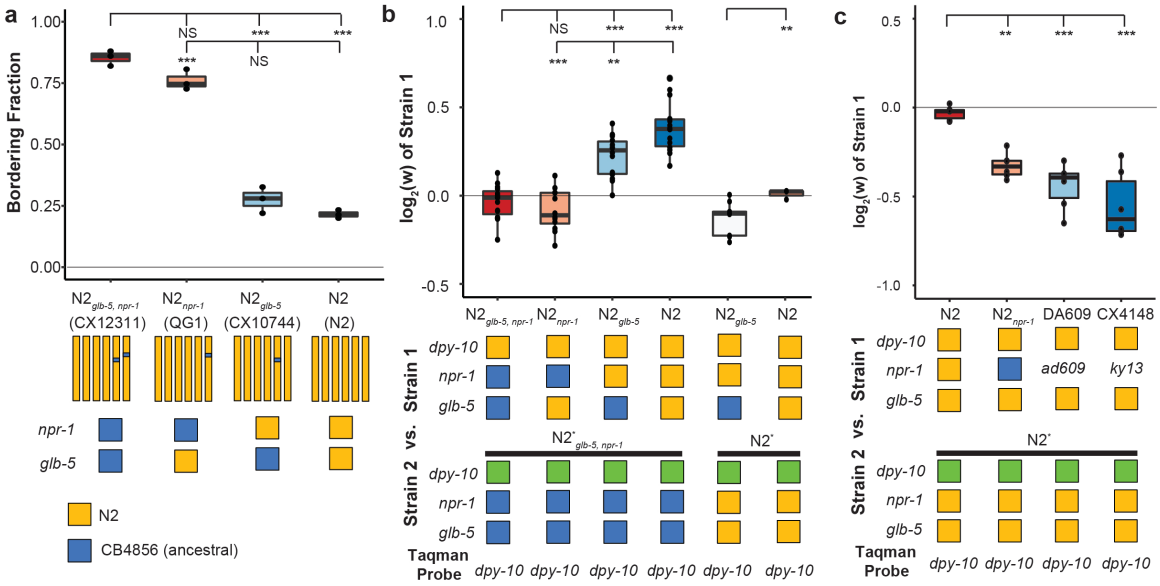

Figure 2.3.2 Derived alleles of *npr-1* and *glb-5* are beneficial. **a**. The bordering rate of the N2 reference strain compared to three near isogenic lines (NILs) containing ancestral alleles of *npr-1* and/or *glb-5* introgressed from the CB4856 wild strain. Bordering rate is defined as the fraction of animals on the edge of the bacterial lawn at a single timepoint. Schematic of each NIL shown below along with the allele of *npr-1* and *glb-5* they contain. Orange represents N2-derived DNA and blue represents CB4856-derived DNA. These strains are referred to by the ancestral alleles they contain (e.g. $N2_{glb-5}$ = CX10744, which is an introgression surrounding *glb-5*). To ascertain statistical significance, ANOVA was used followed by a Tukey's Honest Significant Difference test for multiple comparison tests. NS, not significant, $^{**}p< 0.01$, $^{***}p<0.001$. **b**. Competition experiments between NILs shown in panel **a** against barcoded strains shown in figure **2.3.1c,d**. Green box indicates the strain contains the barcoded allele of *dpy-10*. Positive values indicate Strain

1 is more fit; negative values indicate Strain 2 is more fit. NS not significant, ***p< 0.01, ****p<0.001* by ANOVA with Tukey's Honest Significant Difference test or Wilcoxon-Mann-Whitney nonparametric test. **c.** Competition experiments between strains containing two loss-of-function alleles of *npr-1* (*ad609* and *ky13*) along with controls. ***p< 0.01, ****p<0.001* by ANOVA with Tukey's Honest Significant Difference test.

While the introgressions surrounding the *npr-1* and *glb-5* genes are relatively small, these NIL strains carry additional polymorphisms in surrounding genes from the CB4856 strain. We also performed competition experiments using two previously published *npr-1* loss-of-function alleles *(ad609* and *ky13)* [75] against the N2 barcoded strains. Both the *npr-1(ad609)* and *npr-1(ky13)* loss-of-function alleles decreased the animal's relative fitness in an amount comparable to the ancestral allele (**Figure 2.3.2c**). We did not perform similar experiments on the *glb-5* gene. Altogether, our work suggests that the *npr-1* derived allele increases fitness of animals in laboratory conditions and also suggests that the derived allele of *glb-5* increases the fitness of animals in a *npr-1* dependent manner.

2.3.2 Suppression of social/solitary behavior differences between N2 and CX12311 does not suppress their fitness differences

Animals with reduced function of *npr-1* sense environmental O_2 levels and aerotax towards their preferred O_2 levels (10%) in the presence of foods, which results in aggregation of animals at the borders of the lawn [5,77,111]. This behavior can be suppressed by lowering environmental O_2 levels to the animals preferred O_2 concentrations [77]. We decided to use this environmental manipulation to test the hypothesis that the social foraging behavior

was necessary for the fitness disadvantage experienced by strains containing the ancestral alleles of *npr-1* and *glb-5*. Our above experiments hinted that this hypothesis might be incorrect as the derived *glb-5* allele reduced bordering behavior in the ancestral *npr-1* background without an associated increase in fitness. We first confirmed that we could suppress the bordering behavior differences between $N2_{glb\text{-}5,\ npr\text{-}1}$ and N2 by reducing environmental O_2 levels to 10% or 3% using a Biospherix chamber (**Figure 2.3.3a** and **Videos 1-4**). $N2_{glb\text{-}5,\ npr\text{-}1}$ animals did not form any social groups in the center of the lawn at the lowered O_2 levels and were also indistinguishable from N2 by visual inspection. We also verified that this O_2 manipulation also suppressed roaming/dwelling behavior (**Figure 2.3.3b**). While feeding, *C. elegans* worms alternate between bouts of active exploration (roaming) and periods of inactive movement (dwelling). Animals that are mutant for *npr-1* show increased amounts of roaming behavior [112].

Despite the behavioral similarity of these animals at these lower O_2 levels, the relative fitness differences between the N2 and $N2_{glb\text{-}5,\ npr\text{-}1}$ strains remained (**Figure 2.3.3c**). To further confirm that aggregation behavior was not necessary for the fitness differences, we also performed competition experiments on uniform bacterial lawns (UBLs), which are constructed so that the entire plate is covered with a thin bacterial lawn to remove the O_2 gradients created by the unequal thickness of bacteria in normal lawns. UBLs have been used to suppress *npr-1*-dependent differences in survival in response to bacterial pathogens [113], however, the UBLs were unable to suppress the fitness advantage of N2 animals (**Figure 2.3.3d**).

Animals that carry the ancestral *npr-1* allele can burrow into agar when food is depleted [75], raising the possibility that the fitness gains of N2 could be a result of the transfer process,

which selects for animals on the surface of plates. While visual inspection of the two strains at 10% and 3% did not reveal any obvious differences in burrowing behavior, we also tested the role of burrowing in the fitness differences more rigorously by using modified nematode growth plates that contain agarose that prevents burrowing [100]. The relative fitness differences between N2 and $N2_{glb\text{-}5,\ npr\text{-}1}$ remained unchanged (**Figure 2.3.3d**).

Finally, we tested whether differences in resistance to infection could be responsible for the differences in fitness. The *E. coli* bacterial strain that is used to feed *C. elegans*, OP50, can also infect and kill animals, resulting in a decreased lifespan [114,115]. Both *glb-5* and *npr-1* have been implicated in innate immunity and survival to pathogen exposure [100,113,116,117]. However, the fitness advantage of the N2 strain compared to the $N2_{glb\text{-}5,\ npr\text{-}1}$ strain remained when animals were competed against each other on OP50 bacteria killed by ultraviolet radiation (**Figure 2.3.3e**). The relative fitness on killed OP50 bacteria was slightly decreased, however, this could reflect differences in population demographics, as the killed OP50 supported less overall growth per plate.

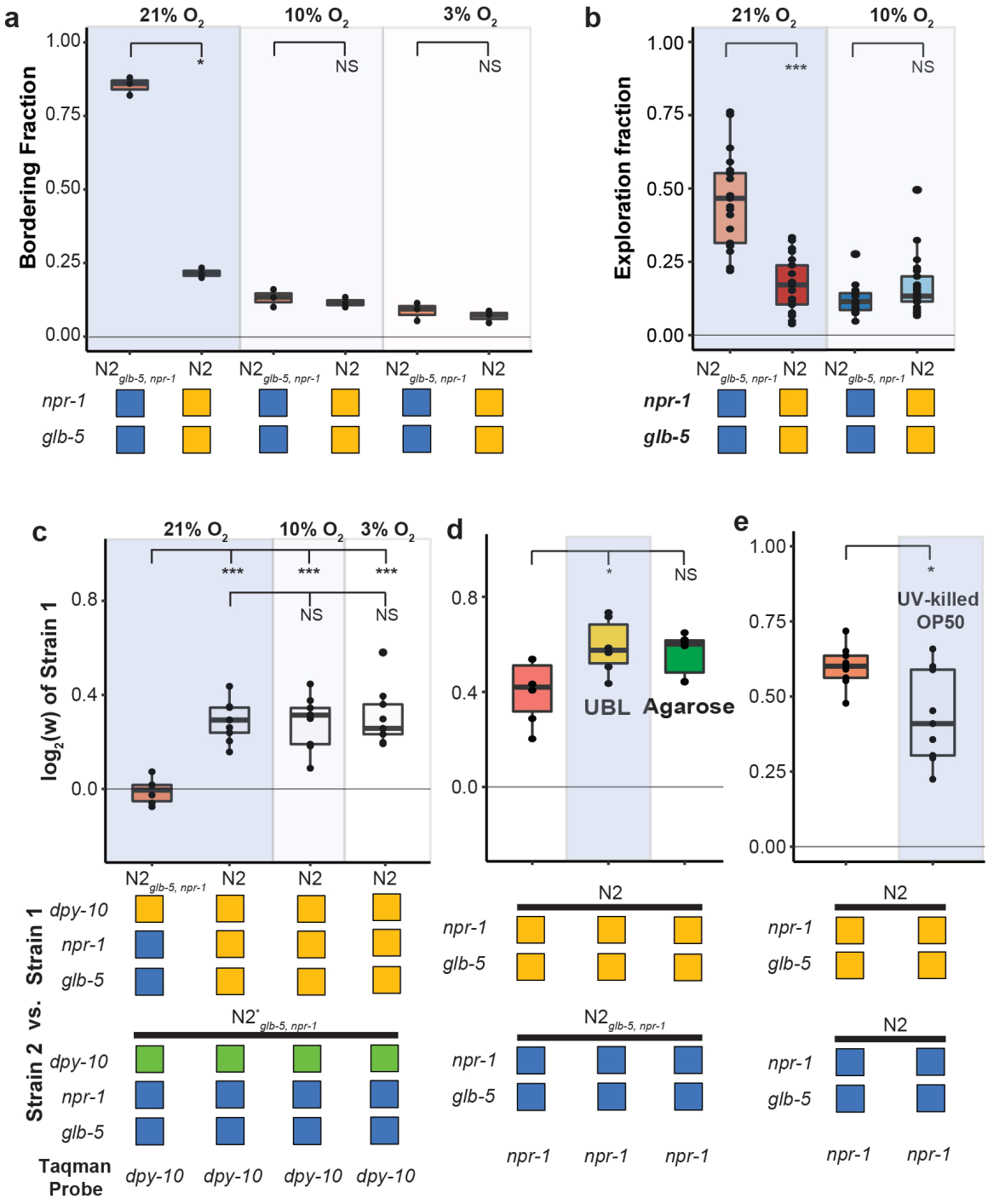

Figure 2.3.3 Fitness advantage of N2 is independent of foraging behavior. **a and b**. Environmental O_2 levels were manipulated using a Biospherix chamber. Differences in (**a**) bordering behavior and (**b**) roaming and dwelling behavior were suppressed in $N2_{glb\text{-}5,\ npr\text{-}}$

$_{1}$ at lower environmental O_2 levels. NS not significant, *$p< 0.05$ by Wilcoxon-Mann-Whitney nonparametric test. **c**. Fitness advantage of N2 over the barcoded $N2_{glb\text{-}5,\ npr\text{-}1}$ strain was independent of environmental O_2. NS, not significant, ***$p<0.001$ by ANOVA with Tukey's Honest Significant Difference test. **d** and **e**. Fitness differences of N2 and $N2_{glb\text{-}5,\ npr\text{-}1}$ on (**d**) uniform bacterial lawns (UBL) where animals were unable to border, on plates containing agarose to prevent burrowing behaviors (NS, not significant, *$p<0.05$ by ANOVA with Tukey's Honest Significant Difference test), and (**e**) on UV-killed bacteria (*$p<0.05$ by Wilcoxon-Mann-Whitney nonparametric test).

These experiments motivated us to also test the relative fitness differences of eleven other wild strains isolated from different parts of the world using strains provided by the *C. elegans* Natural Diversity Resource [118]. Each strain was competed against a barcoded $N2_{glb\text{-}5,\ npr\text{-}1}$. Consistent with their *npr-1* genotype, these wild strains all aggregated at the borders of the bacterial lawn (**Figure 2.3.4a**), but their relative fitness differences varied wildly (**Figure 2.3.4b**). The relative fitness of two of the strains (CB4856 and DL238) was greatly reduced compared to the $N2_{glb\text{-}5,\ npr\text{-}1}$ strain. The relative fitness of five of the strains were comparable to the N2. The relative fitness of the remaining four strains was statistically indistinguishable from the barcoded $N2_{glb\text{-}5,\ npr\text{-}1}$. These results further support that social behavior is not the major determinant of fitness levels in laboratory conditions.

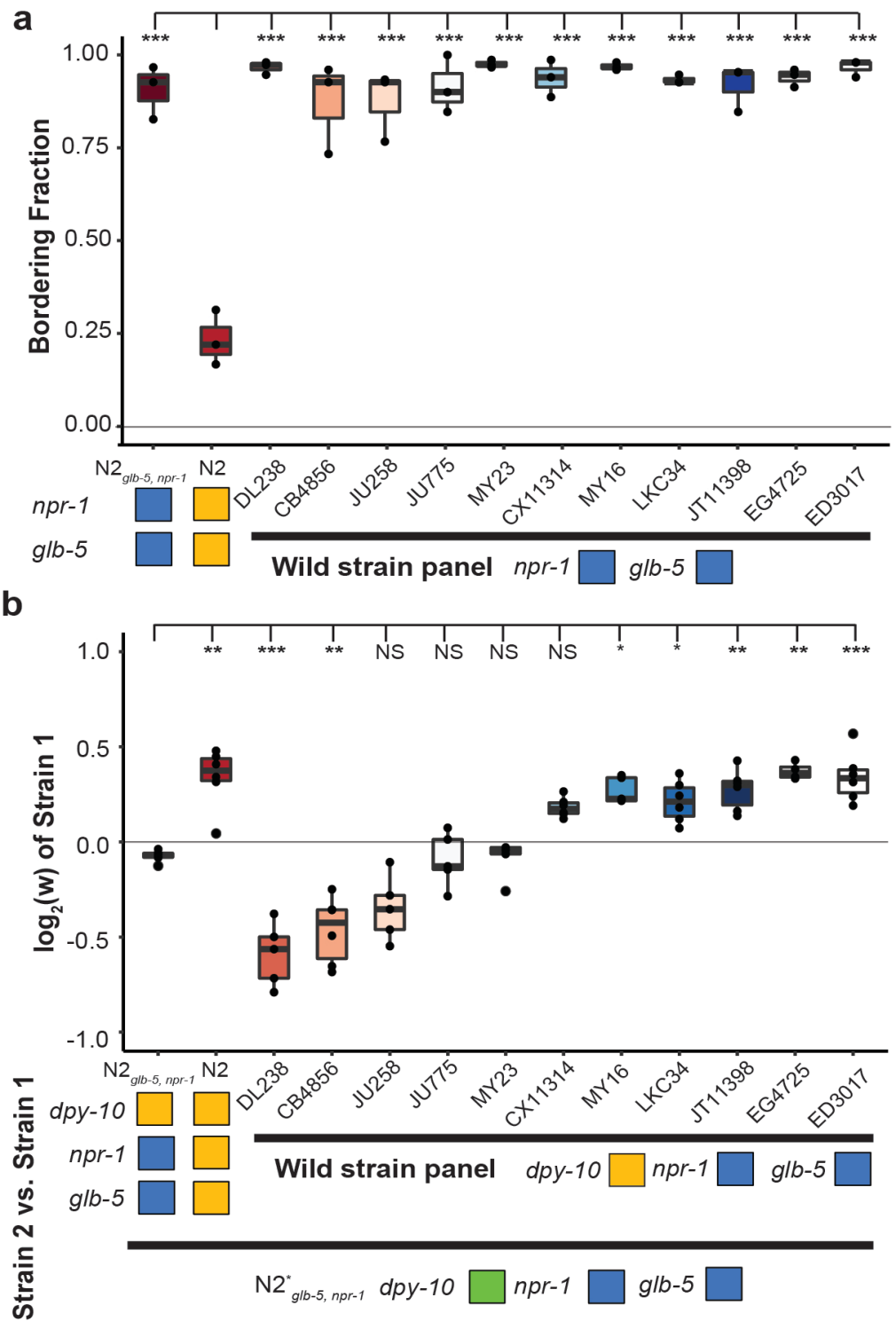

Figure 2.3.4 Bordering rate and relative fitness differences between wild *C. elegans* strains. **a**. A panel of 11 wild strains was tested for bordering behavior. Each of these

wild strains contains ancestral alleles of *glb-5* and *npr-1*. ****p<0.001* by ANOVA with Tukey's Honest Significant Difference test. **b**. Competition experiments between 11 wild strains and barcoded $N2_{glb\text{-}5,\ npr\text{-}1}$ animals. Despite the similarity of bordering behavior, these wild strains displayed a range of relative fitness. NS, not significant, **p<0.05*, ***p<0.01*, ****p<0.001* by ANOVA with Tukey's Honest Significant Difference test.

2.3.3 Development speed and spermatogenesis are increased in N2 in an O_2-independent manner

To gain more insight into the phenotypes that could be responsible for the fitness increases of the N2 strain, we performed RNA sequencing to analyze the transcriptomes of bleach-synchronized N2 and $N2_{glb\text{-}5,\ npr\text{-}1}$ animals grown in either 10% O_2 or 21% ambient O_2 levels. Animals were allowed to develop to the L4 stage and harvested at identical times. We first performed Principal Component Analysis (PCA) analysis on differentially expressed genes to analyze how the environmental and genetic differences globally regulated the transcriptomes of the animals. If environmental O_2 and the genetic background had independent effects on the transcriptomes, we expected to find two major components in the PCA analysis. However, the PCA analysis identified a single component that explained the majority of the variance (77.9%). The genetic and environmental perturbations had similar effects on the first component in an additive manner (**Figure 2.3.5a**). Reducing O_2 levels from 21% to 10% had similar effects on the transcription profiles as changing the background from $N2_{glb\text{-}5,\ npr\text{-}1}$ to N2. Consequently, the animals that differed in both genetic background and environmental O_2 levels (N2 – 21% O_2 vs $N2_{glb\text{-}5,\ npr\text{-}1}$ – 10% O_2) also showed the most similar transcriptional profiles. These patterns were also seen in Hierarchical Clustering using the 1202 differentially-expressed genes

(**Figure 2.3.5b**). These results suggest that the foraging behavioral differences are not responsible for the underlying transcriptomics differences between the different strains and environmental conditions.

The effects of the derived *npr-1* and *glb-5* alleles mimics the effects of lowering environmental O_2 from 21% to 10%. To further gain insight into this connection, we plotted the average transcriptional change between the strain backgrounds vs the average transcriptional change between the environmental O_2 concentrations for each gene (**Figure 2.3.5c**). Surprisingly, we observed a bimodal distribution of values, with a cluster of 652 genes centered at 1.2 $\log_2$-fold change (**Figure 2.3.5c** – red circle). This is unexpected, as it suggests that the environmental and genetic perturbations had identical effects on transcription for all of these genes. When we inspected this list of genes, we noticed a large number of genes that are known to be involved in spermatogenesis. We further investigated the developmental regulation of these 652 genes using previously published transcriptomics data isolated from hermaphrodites or males at specific developmental time points [119] (**Figure 2.3.5d**). The expression of the majority of these genes peaked during the L4 stage in hermaphrodites, was further enriched in L4 males, and suppressed in somatic cells isolated from L4 animals. These observations are consistent with this cluster of genes being involved in spermatogenesis, which occurs during the L4 stage (when RNA was isolated) in hermaphrodite animals.

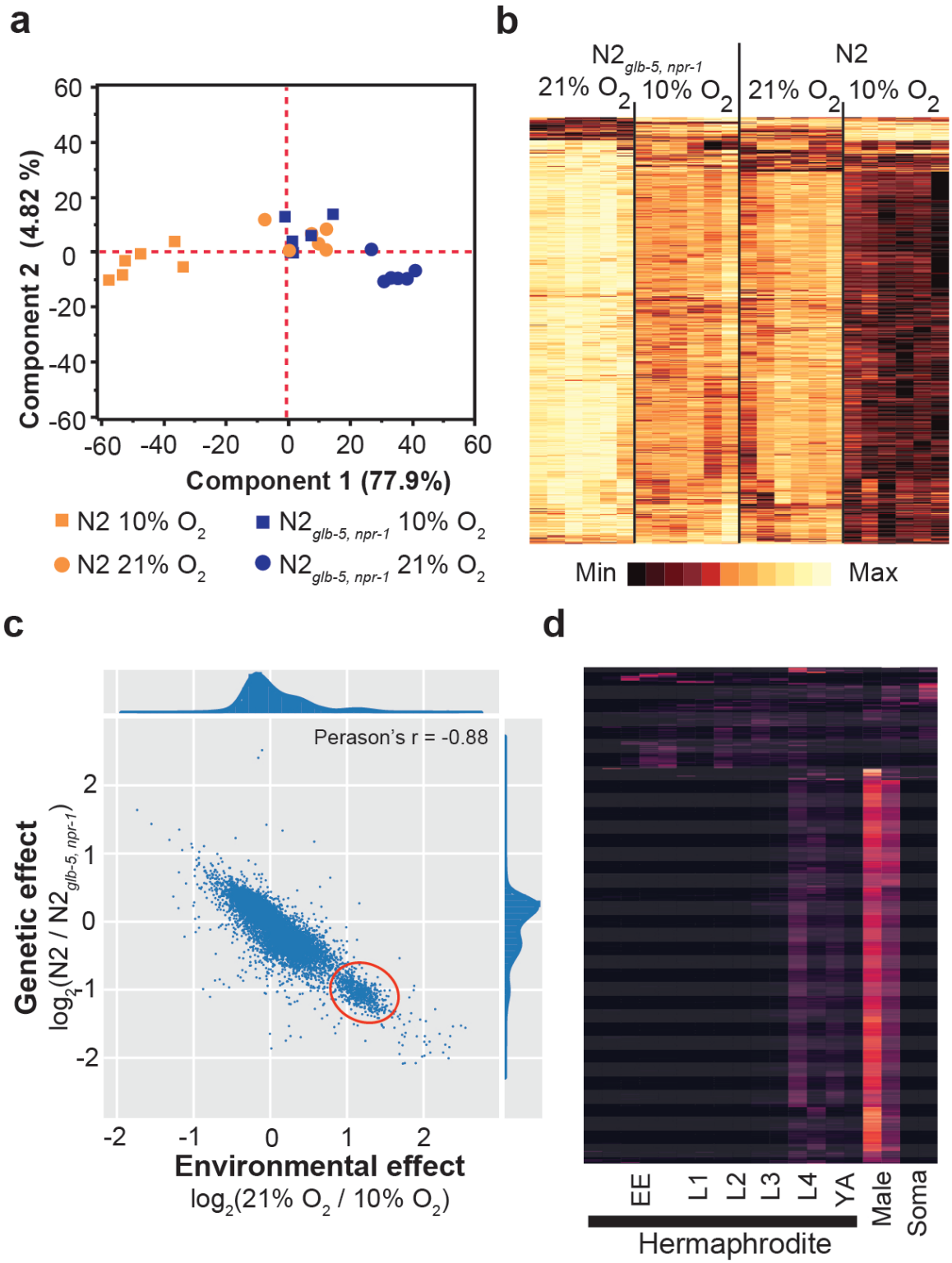

Figure 2.3.5. Differential gene expression analysis of N2 and N2$_{glb\text{-}5,npr\text{-}1}$. a. PCA analysis of transcriptional profiles of bleach-synchronized N2 and N2$_{glb\text{-}5,npr\text{-}1}$ animals

grown in 10% or 21% environmental O_2 (six replicates per strain/condition). The largest two eigenvectors are shown, along with the amount of variance they explain. Developmental age of animals is approximately L4 stage. **b**. Hierarchical clustering of normalized, differentially expressed genes. Columns show strain and conditions; rows show gene expression. **c**. Averaged effect of genotype (y-axis) vs environment (x-axis) for each gene (Supplementary file 1). A small cluster of 652 genes with similar changes is circled in red. **d**. The developmental expression of these 652 genes was further investigated using a previously published dataset. Columns show developmental stage and rows show each gene. Most of these gene peaked in expression in L4 hermaphrodite animals and was further enriched in male L4 animals (Male). Soma indicates expression levels from somatic cells, suggesting this cluster is enriched in germline cells.

We reasoned that the transcriptomics data could indicate a difference in the relative timing of spermatogenesis and/or the number of sperm that are produced in each genetic background/environmental condition. L1 larval stage animals were synchronized; subsequent differences in developmental speed would result in animals in slightly different stages of L4. To test this, we synchronized $N2_{glb\text{-}5,\ npr\text{-}1}$ and N2 animals, placed them in 10% or 21% environmental O_2, and identified the number animals containing mature sperm at 2-hour intervals from 48-56 hours. N2 animals began spermatogenesis approximately two hours earlier than the $N2_{glb\text{-}5,\ npr\text{-}1}$ animals, regardless of the environmental O_2 levels (**Figure 2.3.6a**). Hermaphrodites undergo spermatogenesis for a fixed period of time before permanently switching gametogenesis to the production of oocytes, resulting in the development of a fixed number of self-sperm that are stored in the spermathecae [120]. To

test whether these strains produced the same number of sperm, we used DAPI staining to count the number of sperm found in the spermathecae. Not only did N2 animals start spermatogenesis earlier, they also produced more sperm (**Figure 2.3.6b**). The total fecundity of N2 hermaphrodites that do not mate with males is determined by the number of self-sperm. We confirmed that the difference in self-sperm number also resulted in a larger overall brood size (**Figure 2.3.6c**) and as expected from computational modeling [121], an increased rate of egg-laying later on in life (**Figure 2.3.6d**).

The timing of sexual maturity is an important factor in determining the fitness of animals. We also tested whether the differences in timing of spermatogenesis could lead to differences in when fertilized eggs are produced. We performed similar experiments as above and monitored the time fertilized eggs could be observed in the uterus at two-hour intervals. Again, we observed a difference in N2 and N2$_{glb\text{-}5,\ npr\text{-}1}$ animals at both 10% and 21% environmental O_2 levels. N2 animals were observed to contain fertilized eggs approximately one hour earlier that N2$_{glb\text{-}5,\ npr\text{-}1}$ animals (**Figure 2.3.6e**). The difference in timing of spermatogenesis and fertilization (2 hours vs 1 hour), potentially reflects the fact that N2 animals produce more sperm before switching to oogenesis.

These experiments suggest that the differences in transcription between N2 and N2$_{glb\text{-}5,\ npr\text{-}1}$ could be caused by differences in sexual maturity. We are unable, however, to explain the differences in transcription we observed between 10% and 21% O_2 as mature sperm was observed at similar times in these different environmental conditions (**Figure 2.3.6a**). Potentially the rate of spermatogenesis or expression levels of genes are modified by O_2 levels that are not reflected in the timing of the presence of mature sperm.

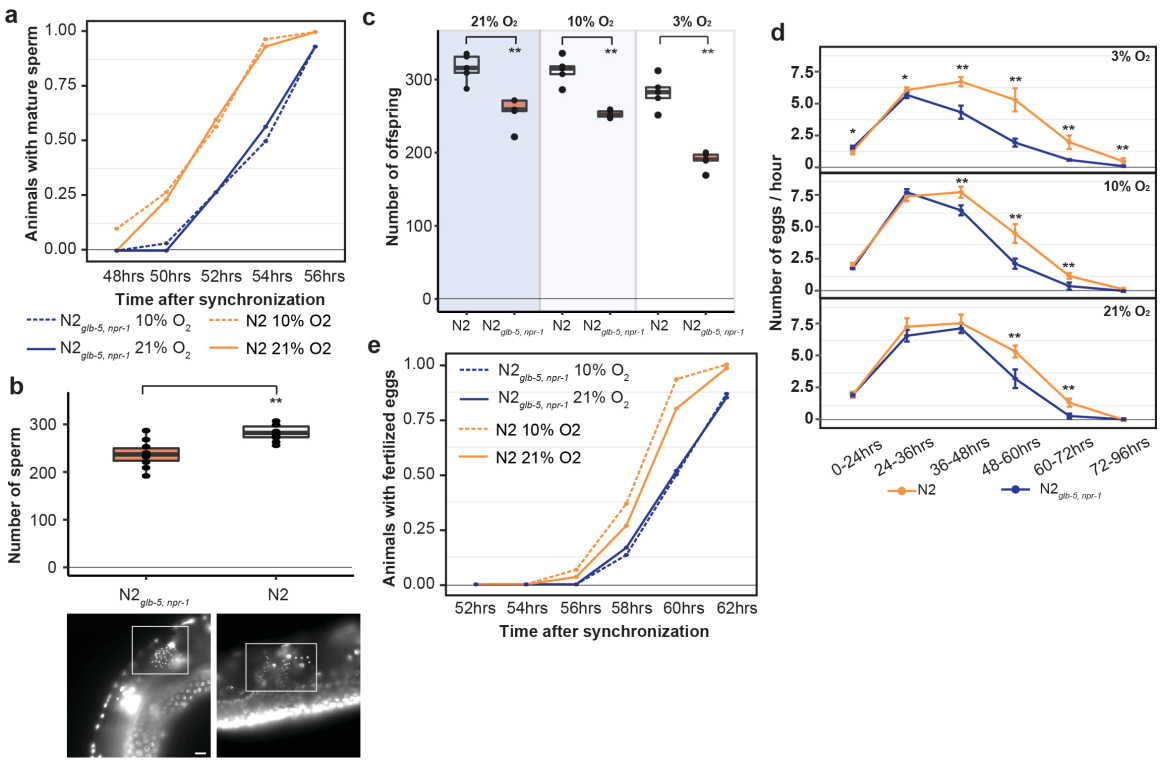

Figure 2.3.6. Reproductive timing in N2 occurs earlier than the N2*glb-5,npr-1* strain. **a**. Animals identified with mature sperm. x-axis indicates time since synchronization using hatch-off. Strain/condition shown in legend. $p = 0.0076$ by Friedman test. **b**. Number of sperm produced by each strain as determined by DAPI straining. Representative images are shown below. Scale bar = 10 µm. ***p< 0.01* by Wilcoxon-Mann-Whitney nonparametric test. **c**. Averaged total number of offspring produced by each strain when grown in different environmental O_2 levels. ***p< 0.01* by Wilcoxon-Mann-Whitney nonparametric test. **d**. Averaged egg-laying rate of L4-synchronized N2 and N2*glb-5,npr-1* animals when grown at different O_2 levels. x-axis indicates time since L4 stage. NS, not significant, **p<0.05*, ***p<0.01* by Wilcoxon-Mann-Whitney nonparametric test. **e**. Number of animals observed with fertilized eggs in their uterus. x-axis indicates time from synchronized egg-lay. $p = 0.0109$ by Friedman test.

2.3.4 *Derived alleles of npr-1 and/or glb-5 increase food consumption in an O_2-independent manner*

Life-history tradeoffs have been proposed in evolutionary theory to account for the linkage between two different traits. Assuming an individual can acquire a finite amount of energy, the investment of energy into one trait leads to consequential changes in other traits as energy resources are shunted into different directions. For example, artificial selection experiments on early fecundity in *C. elegans* resulted in decreased reproduction late in life [122]. The N2 strain seems to violate this tradeoff, as it sexually matures earlier than $N2_{glb\text{-}5,\ npr\text{-}1}$, but also produces more eggs later on in life. We measured the size of N2 and $N2_{glb\text{-}5,\ npr\text{-}1}$ animals and found that N2 animals were also larger than $N2_{glb\text{-}5,\ npr\text{-}1}$ animals at synchronized time points (**Figure 2.3.7a**). These observations suggest that the assumption of a fixed energy acquisition for $N2_{glb\text{-}5,\ npr\text{-}1}$ and N2 might be violated. This would be consistent with Andersen et al's observation that metabolism genes were upregulated by the derived *npr-1* allele, which they proposed represented differences in food intake [100]. It would also be consistent with the role of orthologs of *npr-1* in other species. *npr-1* encodes an ortholog to neuropeptide Y receptors, which are reported to regulate feeding behavior in fishes, birds, and mammals [123-125].

To test this hypothesis, we first utilized a previously described feeding assay to measure the ability of a strain to clear *E. coli* OP50 bacteria from liquid S-media [126]. In this assay, individual wells are seeded with a defined number of bacteria and 20 worms. Each day, the optical density of each well is measured to estimate the amount of food consumed by the worms. In these conditions, N2 cleared the bacteria faster than $N2_{glb\text{-}5,\ npr\text{-}1}$ animals (**Figure 2.3.7b**). While these assays supported our hypothesis, liquid media is fundamentally

different from the conditions experienced on agar plates, making it difficult to generalize the results from one condition to the other. To this end, we developed a new food consumption assay on agar media in 24-well plates. In this assay, each well was seeded with a defined amount of OP50-GFP, which we found could be quantified in a linear manner using a plate reader (**Figure 2.3.7c**). When we tested N2 and $N2_{glb\text{-}5,\ npr\text{-}1}$ animals in 10% or 21% environmental O_2 levels, we found N2 consumed more food than $N2_{glb\text{-}5,\ npr\text{-}1}$ in both environmental conditions (**Figure 2.3.7d**). Interestingly, we found animals grown in 10% O_2 also consume more food than animals grown in 21% O_2. These experiments indicate that N2 animals consume more food than $N2_{glb\text{-}5,\ npr\text{-}1}$.

We next decided to test whether the derived allele of *npr-1* could increase the fitness and feeding rate in a different genetic background. We used the CB4856 wild strain isolated from pineapple fields in Hawaii, which has relatively low relative fitness in laboratory conditions (**Figure 2.3.4b**), taking advantage of a previously constructed NIL of *npr-1* introgressed from N2 into the CB4856 background (CX11400) [108] (**Figure 2.3.7e**). We found that the N2 region surrounding *npr-1* also conferred a fitness advantage in the CB4856 background (**Figure 2.3.7f**). The estimated selective coefficients of the derived allele of *npr-1* was higher in the CB4856 background than the N2 background ($s = 0.61$ vs $s = 0.30$), potentially due to the lower relative fitness of the CB4856 strain. The food consumption of these strains was consistent with the fitness differences (**Figure 2.3.7g**). The derived allele of *npr-1* increased food consumption in both genetic backgrounds but its effect was higher in CB4856.

Food is consumed from the environment by the periodic contraction and relaxation of the pharyngeal muscle which serves to bring material from the environment into the pharynx

and filter out bacterial cells [127]. To test whether the increase in food consumption could be explained by an increase in the rate of pumping, we measured the pharyngeal pumping rate of the $N2_{glb\text{-}5,\ npr\text{-}1}$, N2, CB4856, and CX11400 strains. The effects of the derived allele of *npr-1* was epistatic with respect to the N2 or CB4856 background. The derived allele decreased the pumping rate in the CB4856 background but had no effect in the N2 background (**Figure 2.3.7h**). The effect of the derived allele of *npr-1* on pumping rate is surprising. Pumping rate is often used as a proxy for food consumption; our results indicate that increased pharyngeal pumping does not necessarily lead to increases in food consumption.

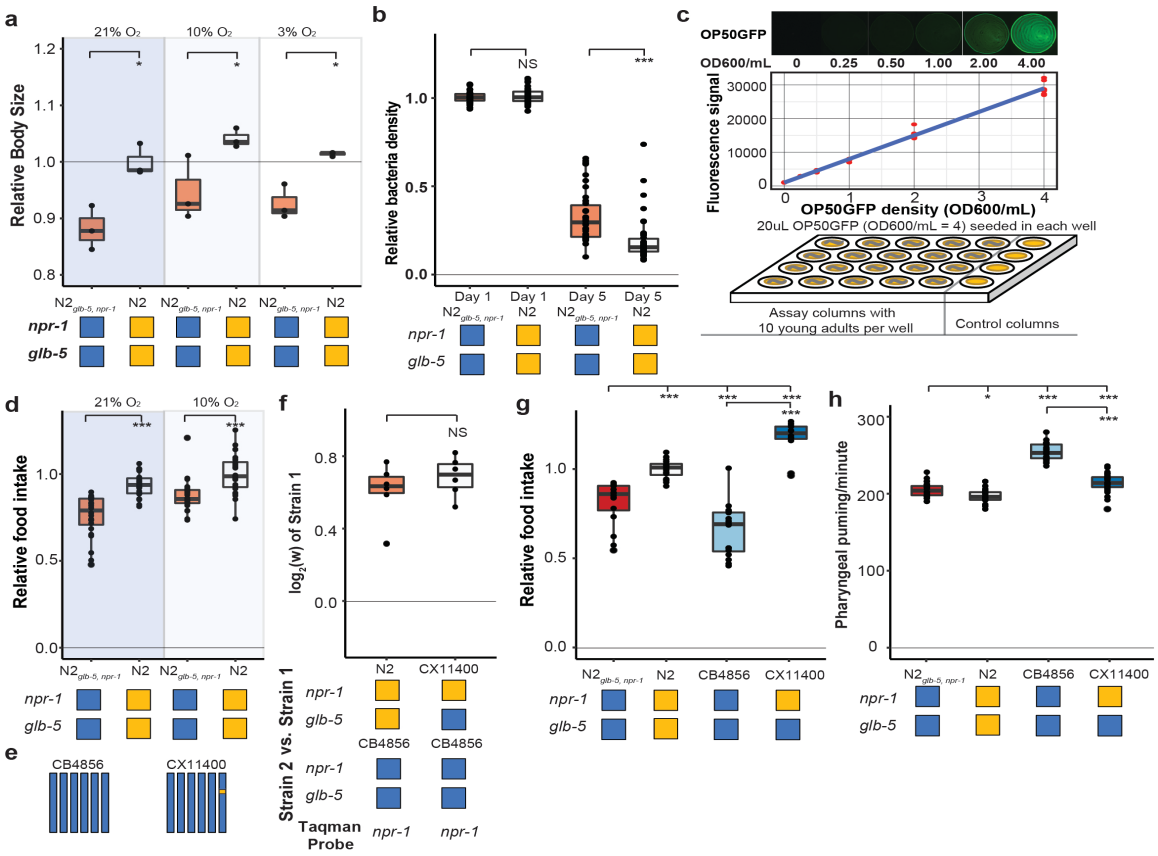

Figure 2.3.7. Derived alleles of *glb-5* and *npr-1* increase food consumption. **a**. N2 and $N2_{glb\text{-}5,npr\text{-}1}$ animals were synchronized by hatch-off and allowed to grow at the indicated

O_2 levels for 72 hours. Video recordings were used to estimate the size of the animals. *$p< 0.05$ by Wilcoxon-Mann-Whitney nonparametric test. **b**. A previously published liquid, bacterial clearing assay was used to estimate food consumption for the $N2_{glb\text{-}5,npr\text{-}1}$ and N2 animals. On day 4, N2 animals had consumed more bacteria than $N2_{glb\text{-}5,npr\text{-}1}$animals. NS, not significant, ***$p<0.001$ by Wilcoxon-Mann-Whitney nonparametric test. **c**. To test food consumption on agar plates, we developed a new assay by seeding 24-well agar plates with defined amounts of OP50-GFP bacteria. The number of bacteria on the plate could be estimated using a microplate reader. **d**. N2 animals consumed more food than $N2_{glb\text{-}5,npr\text{-}1}$ regardless of foraging behaviors. ***$p<0.001$ by Wilcoxon-Mann-Whitney nonparametric test. **e**. Schematic of CB4856 wild strain (blue) and a NIL (CX11400) containing the N2 allele of *npr-1* from N2 (orange). **f.** We tested the fitness effect of the N2 allele of *npr-1* in the CB4856 wild strain using the CX11400 NIL strain. NS, not significant by Wilcoxon-Mann-Whitney nonparametric test. **g**. Food consumption assays between CB4856 and N2 strains or CB4856 and the CX11400 NIL. ***$p<0.001$ by ANOVA with Tukey's Honest Significant Difference test. **h**. Pharyngeal pumping rates of N2, CB4856 and two NIL strains. *$p<0.05$, ***$p<0.001$ by ANOVA with Tukey's Honest Significant Difference test.

We also measured a number of size parameters of the pharynx but found no obvious differences that could account for the increased food consumption (**Figure 2.3.8**). Potentially the pharynx is more efficient at bringing food in from the external environment due to stronger pump strength, more efficient filtering processes or other unknown behavioral differences that contribute to food intake.

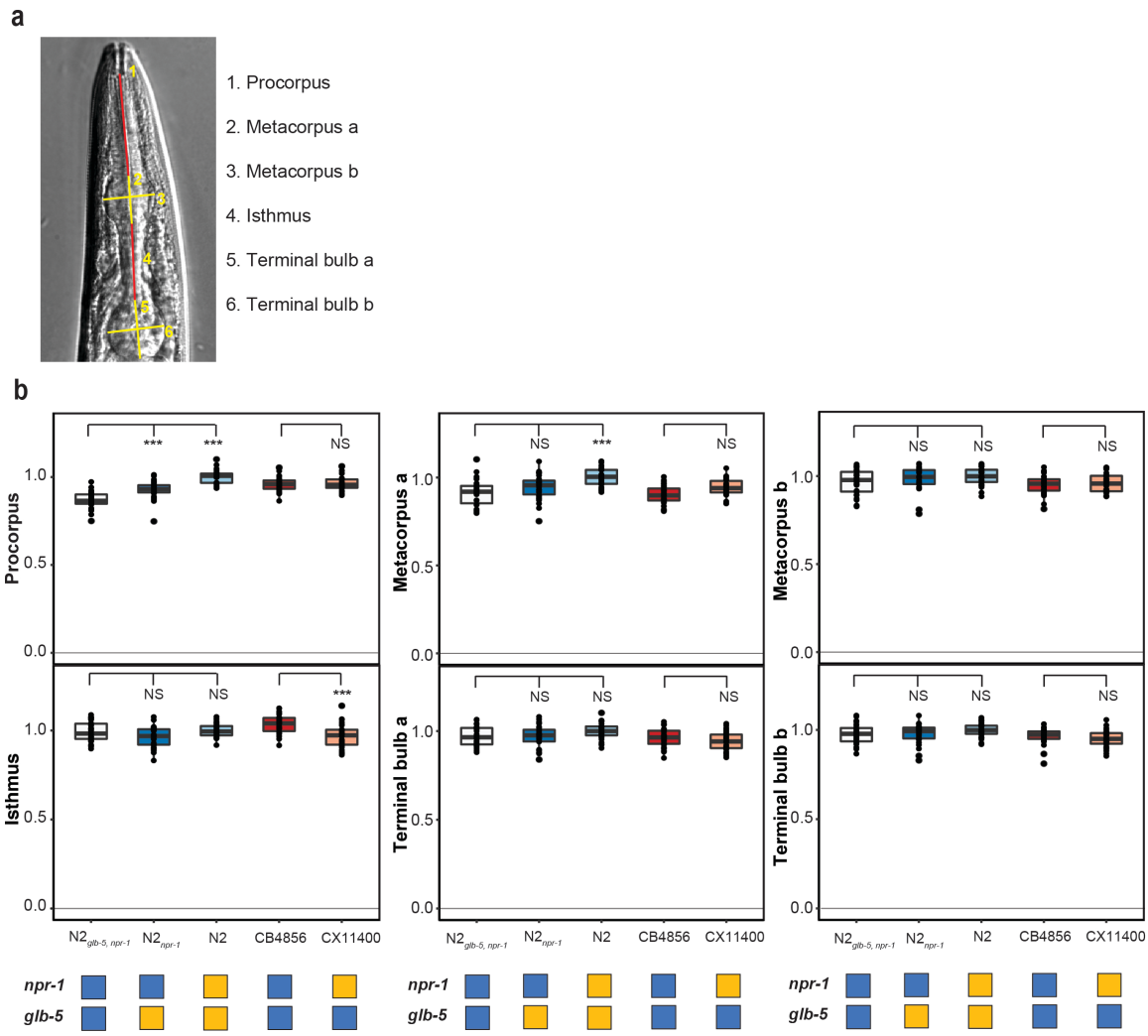

Figure 2.3.8. Measurement of pharyngeal sizes of adult animals. a. Schematic of measurements made for the pharynx. **b.** Pharyngeal sizes of adult animals, normalized to the N2 strain. ****p<0.001* by ANOVA with Tukey's Honest Significant Difference test.

2.3.5 Fitness gains of the derived alleles require the URX, AQR, and/or PQR neurons

We next decided to gain insight into the cellular mechanisms by which *npr-1* and *glb-5* increased fitness of the strains. Previous studies have shown that *npr-1* and *glb-5* regulate

social behavior through the URX-RMG neuronal circuit (**Figure 2.3.9a**). *glb-5* tunes O_2-sensititivies of the URX oxygen-sensing neuron pair through regulation of O_2-sensing guanylyl cyclases, leading to changes in influx of Ca^{++} into the cell body [19,76,128-130]. The derived allele of *npr-1* inhibits the activity of the RMG hub interneuron which suppresses aerotaxis and social behavior [78,131]. The RMG neurons connect to URX and a number of other sensory neurons through gap junctions, which are necessary for foraging behaviors [132]. URX neurons also integrate O_2 with internal nutrient reserves [80]. To test the role of URX in the fitness gains of the *npr-1* and *glb-5* derived alleles, we used the *qaIs2241* integrated cassette that specifically kills the O_2-sensing neurons URX, AQR and PQR [5]. We crossed this cassette into the $N2_{npr-1}$, $N2_{glb-5}$ and $N2_{glb-5,\ npr-1}$ strains and repeated the pairwise competition experiments performed in **Figure 2.3.2a** using strains that now also contained the *qaIs2241* cassette. In all cases, the relative fitness gains of the derived alleles were decreased by the presence of the neuronal ablation (**Figure 2.3.9b**).

These experiments suggest that the derived alleles either activate or disinhibit the URX, AQR, and or PQR neurons which leads to increases in fitness. To distinguish between these possibilities, we competed N2 and $N2_{glb-5,\ npr-1}$ strains with and without the *qaIs2241* against each other. Strains that carried the *qaIs2241* cassette were dramatically less fit than the control worms, suggesting that URX, AQR, and PQR promote fitness in laboratory conditions (**Figure 2.3.9c**).

We and others have shown that *glb-5* and *npr-1* are pleiotropic, regulating social behavior and food consumption. Potentially this pleiotropy arises from the ability of the URX, AQR, and PQR neurons to these biological traits. To test this, we phenotyped strains that carried the *qaIs2241* cassette for social behaviors, food consumption and reproductive timing

(**Figures 2.3.9d-g**). These experiments indicated that these neurons are required for each of these three traits. Interestingly, food consumption in the *qaIs2241* strains was reduced without a corresponding change in pharyngeal pumping rate, further confirming that these phenotypes could be separated from each other at a genetic and cellular level.

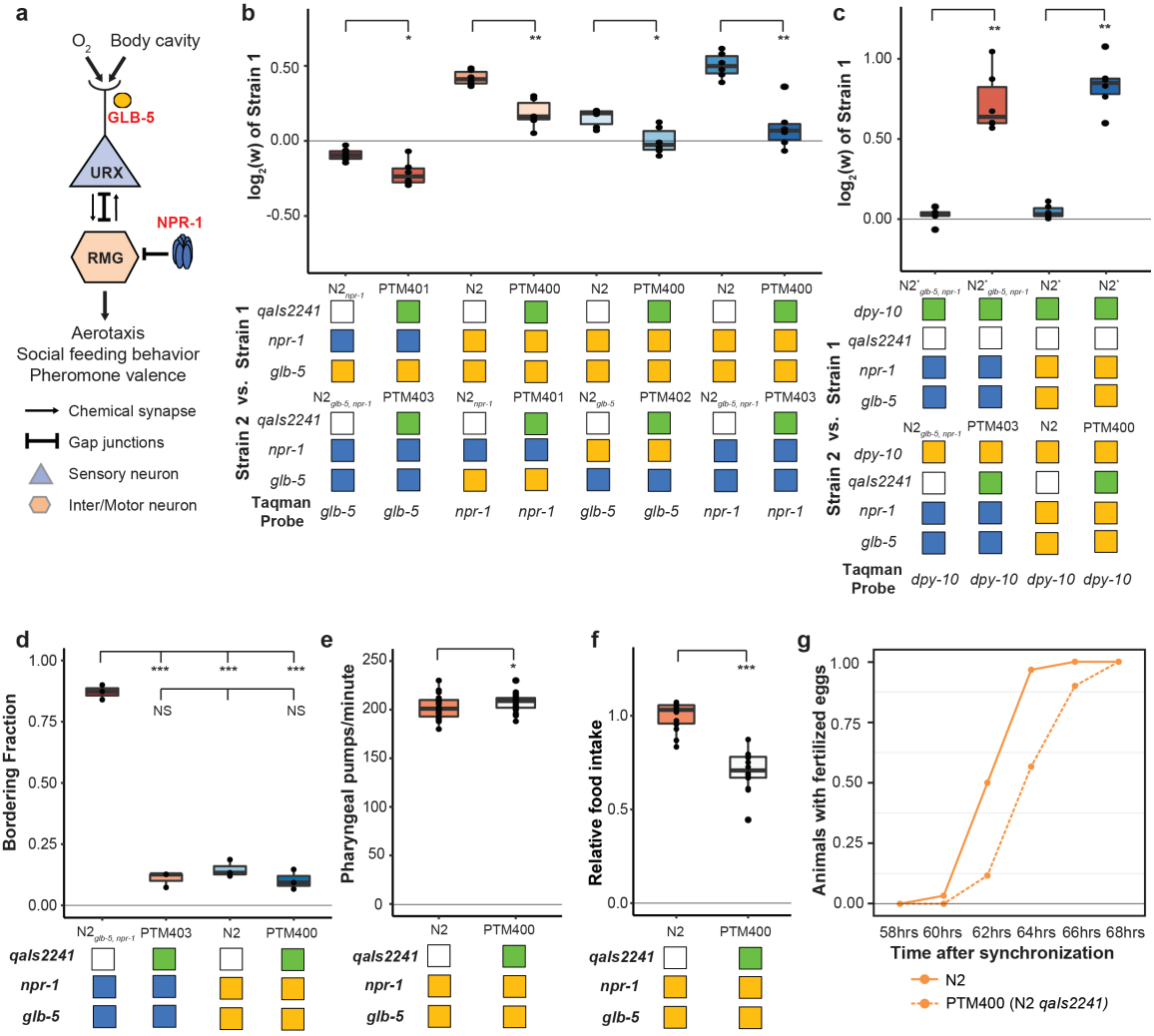

Figure 2.3.9. **O_2-sensing neurons contribute to fitness differences of N2 and N2$_{glb\text{-}5,npr\text{-}1}$.** **a**. Schematic showing putative cellular sites of action for *glb-5* and *npr-1*. *glb-5* modulates O_2 responses in the URX body cavity neurons. *npr-1* is thought to modulate electrical signaling in the RMG hub-and-spoke neuron which forms gap junctions onto

URX. **b.** and **c**. Competition experiments between indicated strains. *qaIs2241* is an integrated genetic cassette that ablates the URX, AQR, and PQR neurons. Green indicates the presence of the cassette (and loss of URX, AQR, and PQR neurons). *$p< 0.05$, **$p < 0.01$ by Wilcoxon-Mann-Whitney nonparametric test. **d**. Bordering rates of indicated strains. The *qaIs2241* cassette suppresses bordering of the $N2_{glb\text{-}5,npr\text{-}1}$ strains. NS not significant, *$p< 0.05$ by Wilcoxon-Mann-Whitney nonparametric test. **e**. Pharyngeal pumping rates of N2, and N2 strains carrying the *qaIs2241* cassette. *$p<0.05$ by Wilcoxon-Mann-Whitney nonparametric test. **f**. Relative food consumption rates between the indicated strains. ***$p<0.001$ by Wilcoxon-Mann-Whitney nonparametric test. **g**. Number of animals observed with fertilized eggs in their uterus. x-axis indicates time from synchronized egg-lay. $p = 0.0455$ by Friedman test.

2.3.6 Fitness gains, increased food consumption, and earlier reproductive timing in N2 require the daf-22 gene

We also decided to test whether ascaroside pheromones were necessary for the fitness differences between N2 and $N2_{glb\text{-}5,npr\text{-}1}$. Nematodes release a number of ascaroside molecules, which are in turn sensed by a distributed neural circuit that integrates and modifies a number of behavioral and developmental phenotypes [133,134]. There are a few reasons to think that ascaroside pheromones might be involved in the fitness gains of the N2 strain. First, work by Andersen et. al indicated that population density directly impacts lifetime fecundity and adult body length differences between N2 and CB4856 strains [100]. Second, our previous studies of *C. elegans* domestication to liquid cultures has found that

pheromone signaling was modified by fixed genetic changes [71,105]. Finally, the derived alleles of *npr-1* and *glb-5* have been shown to modify pheromone valence in a variety of contexts [78,130,135,136]. To test the role of ascaroside pheromones, we followed previous publications using a genetic knockout of the *daf-22* gene, which encodes a peroxisomal enzyme required for the biosynthesis of *C. elegans* pheromones [137] and accumulation of lipid droplets [138], using CRISPR-Cas9 enabled genome editing to create a large deletion of *daf-22* in the N2 strain, which was then crossed to the $N2_{glb\text{-}5,npr\text{-}1}$background. Competition experiments demonstrated that *daf-22* was necessary for the fitness advantage of derived *npr-1* and *glb-5* alleles (**Figure 2.3.10a**). In addition, *daf-22* was necessary for the faster sexual maturity (**Figure 2.3.10b**) and increased food intake (**Figure 2.3.10c**) of the N2 strain compared to $N2_{glb\text{-}5,npr\text{-}1}$. These data suggest that *npr-1* and *glb-5* reprogram pheromone responses resulting in increased sexual maturity and ability to consume food.

daf-22 encodes a peroxisomal fatty acid ß-oxidation enzyme. Besides its role in the biosynthesis of ascaroside pheromones, *daf-22* has recently been shown to play a distinct role in ASK neurons, where it is required for the metabolization of fatty acids that stimulate the endoplasmic reticulum stress response, promoting the transcription of insulin-like peptides that regulate dauer formation and other biological processes [139]. *daf-22* mutants also accumulate massive amounts of fatty acids and fatty acyl-CoAs in their intestines [137,140,141], which can potentially regulate feeding behavior through homeostasis mechanisms [142]. To determine if the differences observed in the *daf-22* mutants were caused by the lack of ascaroside pheromones, we attempted to rescue these phenotypes using two concentrations of crudely purified pheromones isolated from animals grown in liquid cultures. Neither of these concentrations were able to rescue the differences in food

consumption or reproductive timing (**Figure 2.3.10d, e**). These experiments suggest that the effects of the *daf-22* mutants we have observed might be independent of their role in producing ascaroside pheromones.

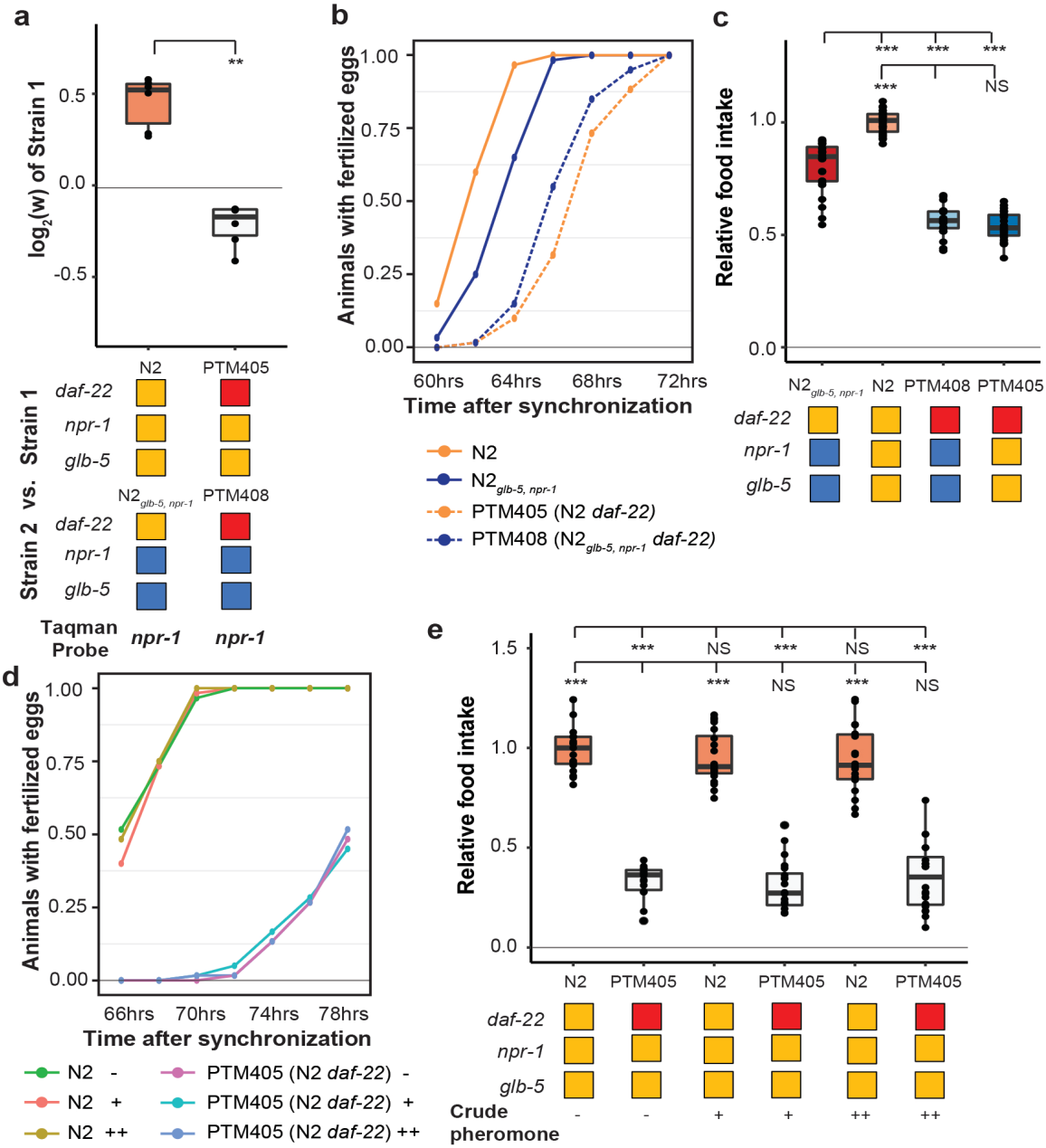

Figure 2.3.10. ***daf-22*** **is required for fitness differences of N2 and N2*glb-5,npr-1*** **a**. Competition experiments between indicated strains. *daf-22* encodes a sterol carrier protein,

which is required for biosynthesis of most ascaroside pheromones. Red indicates the strain contain a deletion that spans the gene. $**p < 0.01$ by Wilcoxon-Mann-Whitney nonparametric test. **b**. Number of animals that carry fertilized eggs at the indicated timepoints. $p = 6.61 \times 10^{-4}$ by Friedman test. **c.** On plate feeding assays of the indicated strains. NS, not significant, $***p<0.001$ by ANOVA with Tukey's Honest Significant Difference test. **d** and **e**. Attempts to rescue the relative food intake and reproductive timing defects of the *daf-22* strain using crude pheromone. Neither of two concentrations of crude pheromone isolated from animals grown in liquid cultures had a significant effect on the two traits. NS, not significant, $***p<0.001$ by ANOVA with Tukey's Honest Significant Difference test. $p = 7.45 \times 10^{-6}$ by Friedman test.

2.4 Discussion

In this report, we studied the fitness consequences of two derived alleles that arose and fixed in the N2 strain after isolation from the wild. We find that both alleles can be adaptive, with selective coefficients that are larger than many characterized beneficial alleles from other species. These results are consistent with the derived alleles spreading through the ancestral N2 populations due to positive selection. If this was true, it would suggest that the derived allele of *npr-1* arose first, as the derived *glb-5* allele is only beneficial in this derived genetic background. However, the demographic history and laboratory environment of how N2 was grown at the time these alleles arose is largely lost [70]. The exact laboratory growth conditions (liquid axenic vs. solid media), transfer processes (picking vs. chunking) and effective population sizes (between 4 and 1000) used to

propagate a *C. elegans* strain is incredibly variable. It is likely that the evolutionary forces responsible for the fixation of these alleles will remain lost to history.

Nevertheless, the ability of positive selection to act upon the derived *npr-1* allele can be observed in current experiments. A recent example is provided by Noble and colleagues, who created a large mapping population between sixteen parental strains (including N2 and CB4856) to create a large panel recombinant inbred lines (RILs) [102]. During the outcrossing phase of construction, the N2 allele of *npr-1* spread through the population to fixation, consistent with its dominant action and the strong selective advantage of this allele. Potentially, variation in *npr-1* affected allele frequencies of unlinked loci as well. For example, an excess of CB4856 haplotypes was observed in the RILs, suggesting that CB4856 haplotypes were more likely to contain beneficial alleles. Our measurements of the relative fitness of the CB4856 strain, however, creates an apparent paradox, as CB4856 was one of the least fit strains among the wild strains we tested (**Figure 2.3.4b**). Potentially, epistatic interactions between CB4856 alleles and the derived allele of *npr-1* could help resolve this; the effect of *npr-1* on food intake and fitness is higher in the CB4856 background (**Figure 2.3.7f,g**). Differences in effect size of a focal allele in different genetic backgrounds is considered evidence for the existence of epistasis [143]. Potentially, the presence of laboratory-derived alleles in mapping populations will skew not only the allele frequencies of these beneficial alleles, but also natural genetic variants that interact epistatically with them.

Evolution of behavioral traits is one strategy for animals to respond to a new environment. The identification of a polymorphism in *npr-1* has served as an example of how behavioral variation can arise from genetic variation. However, our work suggests that the

social/solitary feeding behavioral changes of N2 are not sufficient for explaining its fitness gains. Rather, we propose that changes to food intake, sexual maturity, and fecundity are more important. One unresolved question is why wild strains do not eat as much food as the N2 strain. We believe there must be some sort of tradeoff – either energetically or developmentally – that makes the derived mutation unfavorable in their natural environments. Mechanistic understanding of the energetic forces necessary for *C. elegans* to bring food into their pharynx is lacking. In fact, pharyngeal pumping rates are often used as proxies for food intake, which we have shown here can be unrelated to the amount of food consumed. Potentially the thick slurry of food in laboratory plates is completely different biophysically from the mixed bacterial species encountered on rotting material in the wild. Alternatively, differences in feeding behaviors unrelated to social/solitary behaviors might also mediate the differences in food intake. Our experiments suggest that previously described roaming/dwelling differences are also not responsible, however, additional uncharacterized behavioral differences could influence food intake.

The changes to fitness and food consumption in the N2 strain appear to be mediated by the nervous system, which we propose occurs through changes in the function and/or downstream effects of the URX sensory neurons. In this paper, we have also shown that animals that lack the URX neurons consume food at lower rates. How does URX modify food consumption? One possibility is URX regulates pharyngeal neurons extrasynaptically through neuropeptides or through chemical synapses onto the RIP interneurons, which represent the only connection between the somatic and pharyngeal nervous system. Alternatively, URX could regulate food consumption indirectly by stimulating metabolism of fatty acids. URX, along with AQR and PQR, are body cavity neurons, sending ciliated

dendrites into the coelomic fluid, which serves as the circulatory system for *C. elegans* [144]. Besides sensing external O_2, URX neurons monitor fat stores, which are thought to regulate tonic Ca^{++} responses of the URX neurons [80]. URX, in turn, can stimulate fat loss, creating a homeostatic loop that ensures that fat mobilization only occurs when there are sufficient fat reserves and when environmental O_2 is high enough to metabolize the fatty acids into energy [80]. The changes to the N2 strain could have resulted in URX triggering fatty acid metabolism at a higher rate in laboratory conditions. The access energy could be used to speed development and increase growth. Why, then, would the animals consume more food? Fat metabolism has been shown to regulate satiety behavior in *C. elegans*, which could account for increases in food consumption we see in these strains [142].

This model could also explain the effects we see in the *daf-22* mutants, which we originally explored for the potential role of pheromone responses in feeding and fitness changes. *daf-22* mutants accumulate large amounts of fatty acids in their bodies, which potentially inhibits the food consumption rates of these animals. However, our experiments do not preclude a role for pheromones in these fitness and food consumption changes to N2. Potentially our crude purifications do not capture physiologically relevant levels and ratios of the complex pheromone mixtures. Pheromones might also contribute to these differences in combination with other *daf-22* dependent pathways. Primer pheromones have been shown to influence body fat metabolism in *C. elegans* through the ADL sensory neuron [145]. ADL sensory neurons are regulated by pheromones in an *npr-1*-dependent manner [78,128,132,135,136]. It is possible these changes, or other parts of the pheromone circuits are also necessary. Future experiments, enabled by the development of the on-plate food consumption assay, should be enlightening.

Our work underscores issues with growing organisms in the laboratory for multiple generations. Despite the attempts of researchers to create fertile conditions for nematodes to grow in, we found a large difference in relative fitness between different strains of *C. elegans* when competed in the laboratory. Natural genetic variation and *de novo* variation both result in fitness differences that selection can act on. Experimenters using wild strains of nematodes must take care in designing experiments to account for this, especially in wild strains with lower initial fitness levels. We believe that the laboratory selection pressures we characterized here will generalize to other invertebrate and vertebrate animals. If so, the behaviors and physiology of these animals will also be modified over generations of growth. Our work suggests that not only will the traits that confer fitness advantages be modified, but potentially additional traits due to the pleiotropic actions of many genes, and relaxed stabilizing selection on traits in laboratory conditions.

2.5 Materials and Methods

Table 1 - Key resources for Chapter 2

Reagent type or resource	Designation	Source of reference	Identifiers	Additional information
Gene (*C. elegans*)	*npr-1*	Worm base	Wormbase ID: WBGene00003807	Sequence: C39E6.6
Gene (*C. elegans*)	*glb-5*	Worm base	Wormbase ID: WBGene00015964	Sequence: C18C4.1
Gene (*C. elegans*)	*dpy-10*	Worm base	Wormbase ID: WBGene00001072	Sequence: T14B4.7
Gene (*C. elegans*)	*daf-22*	Worm base	Wormbase ID: WBGene00013284	Sequence: Y57A10C.6
strain, strain background (*E. coli*)	OP50	Caenorhabditis genetics center (CGC)	RRID:WB-STRAIN:OP50	

Table 1 – continued

Reagent type or resource	Designation	Source of reference	Identifiers	Additional information
strain, strain background (*E. coli*)	OP50 GFP	Caenorhabditis genetics center (CGC)	RRID:WB-STRAIN:OP50-GFP	with pFPV25.1 express GFP.
Strain (*C. elegans*)	N2	Cori Bargmann Lab (The Rockefeller University)	RRID:WB-STRAIN:N2	
Strain (*C. elegans*)	CB4856	*Caenorhabditis elegans* Natural Diversity Resource (CeNDR)	RRID:WB-STRAIN:CB4856	Website: https://www.elegansvariation.org/
Strain (*C. elegans*)	DL238	*Caenorhabditis elegans* Natural Diversity Resource (CeNDR)	RRID:WB-STRAIN:DL238	Website: https://www.elegansvariation.org/
Strain (*C. elegans*)	JU258	*Caenorhabditis elegans* Natural Diversity Resource (CeNDR)	RRID:WB-STRAIN:JU258	Website: https://www.elegansvariation.org/
Strain (*C. elegans*)	JU775	*Caenorhabditis elegans* Natural Diversity Resource (CeNDR)	RRID:WB-STRAIN:JU775	Website: https://www.elegansvariation.org/
Strain (*C. elegans*)	MY16	*Caenorhabditis elegans* Natural Diversity Resource (CeNDR)	RRID:WB-STRAIN:MY16	Website: https://www.elegansvariation.org/
Strain (*C. elegans*)	MY23	*Caenorhabditis elegans* Natural Diversity Resource (CeNDR)	RRID:WB-STRAIN:MY23	Website: https://www.elegansvariation.org/
Strain (*C. elegans*)	CX11314	*Caenorhabditis elegans* Natural Diversity Resource (CeNDR)	RRID:WB-STRAIN:CX11314	Website: https://www.elegansvariation.org/
Strain (*C. elegans*)	LKC34	*Caenorhabditis elegans* Natural Diversity Resource (CeNDR)	RRID:WB-STRAIN:LKC34	Website: https://www.elegansvariation.org/
Strain (*C. elegans*)	ED3017	*Caenorhabditis elegans* Natural Diversity Resource (CeNDR)	RRID:WB-STRAIN:ED3017	Website: https://www.elegansvariation.org/
Strain (*C. elegans*)	JT11398	*Caenorhabditis elegans* Natural Diversity Resource (CeNDR)	RRID:WB-STRAIN:JT11398	Website: https://www.elegansvariation.org/
Strain (*C. elegans*)	EG4725	*Caenorhabditis elegans* Natural Diversity Resource (CeNDR)	RRID:WB-STRAIN:EG4725	Website: https://www.elegansvariation.org/

Table 1 – continued

Reagent type or resource	Designation	Source of reference	Identifiers	Additional information
Strain (*C. elegans*)	PTM229	This paper	RRID:WB-STRAIN:PTM229	Strain Background: N2
Strain (*C. elegans*)	PTM288	This paper	RRID:WB-STRAIN:PTM288	Strain Background: N2
Strain (*C. elegans*)	PTM289	This paper	RRID:WB-STRAIN:PTM289	Strain Background: N2
Strain (*C. elegans*)	PTM95	PMID: 27467070	RRID:WB-STRAIN:PTM95	Strain Background: N2
Strain (*C. elegans*)	CX12311	PMID: 21849976	RRID:WB-STRAIN:CX12311	Strain Background: N2
Strain (*C. elegans*)	QG1	PMID: 27172189	RRID:WB-STRAIN:QG1	Strain Background: N2
Strain (*C. elegans*)	CX10774	PMID: 19285466	RRID:WB-STRAIN:CX10774	Strain Background: N2
Strain (*C. elegans*)	CX11400	PMID: 23284308	RRID:WB-STRAIN:CX11400	Strain Background: CB4856
Strain (*C. elegans*)	CX4148	PMID: 9741632	RRID:WB-STRAIN:CX4148	Strain Background: N2
Strain (*C. elegans*)	DA609	PMID: 9741632	RRID:WB-STRAIN:DA609	Strain Background: N2
Strain (*C. elegans*)	CX7102	PMID: 16903785	RRID:WB-STRAIN:CX7102	Strain Background: N2
Strain (*C. elegans*)	PTM400	This paper	RRID:WB-STRAIN:PTM400	Strain Background: N2

Table 1 – continued

Reagent type or resource	Designation	Source of reference	Identifiers	Additional information
Strain (*C. elegans*)	PTM400	This paper	RRID:WB-STRAIN:PTM400	Strain Background: N2
Strain (*C. elegans*)	PTM401	This paper	RRID:WB-STRAIN:PTM401	Strain Background: N2
Strain (*C. elegans*)	PTM402	This paper	RRID:WB-STRAIN:PTM402	Strain Background: N2
Strain (*C. elegans*)	PTM403	This paper	RRID:WB-STRAIN:PTM403	Strain Background: N2
Strain (*C. elegans*)	PTM404	This paper	RRID:WB-STRAIN:PTM404	Strain Background: N2
Strain (*C. elegans*)	PTM405	This paper	RRID:WB-STRAIN:PTM405	Strain Background: N2
Strain (*C. elegans*)	PTM408	This paper	RRID:WB-STRAIN:PTM408	Strain Background: N2
Sequence-based reagents (Plasmid)	Plasmid: pDD162 PrU6::*dpy-10_sgRNA*	PMID: 27467070		CRISPR/Cas9 gene editing sgRNA
Sequence-based reagents (Plasmid)	Plasmid: pDD162 Pr*eft3::Cas9*	PMID: 27467070		CRISPR/Cas9 gene editing Cas9
Sequence-based reagents (Oligonucleotide)	*dpy-10 (cn64)*	PMID: 25161212		CRISPR/Cas9 gene editing DNA repair oligo for inducing *dpy-10 cn64* mutation
Sequence-based reagents (Oligonucleotide)	*dpy-10 (kah82/kah83)*	This paper		CRISPR/Cas9 gene editing DNA repair oligo for inducing *dpy-10* Thr90 slient mutation
Sequence-based reagents (Oligonucleotide)	*dpy-10 (kah84)*	This paper		CRISPR/Cas9 gene editing DNA repair oligo for inducing *dpy-10* Arg92 slient mutation

Table 1 – continued

Reagent type or resource	Designation	Source of reference	Identifiers	Additional information
Chemical compound, drug	1x Antibiotic-Antimycotic	ThermoFisher	Cat. No.: 15240062	
Chemical compound, drug	FUDR	Sigma	Cat. No.: F0503	
Commercial assay, kit	Taqman probe: *dpy-10 (kah82/kah83)*	ThermoFisher: Custom TaqMan® SNP Genotyping Assays	PTM09	
Commercial assay, kit	Taqman probe: *dpy-10 (kah84)*	ThermoFisher: Custom TaqMan® SNP Genotyping Assays	PTM10	
Commercial assay, kit	Taqman probe: *npr-1(g320)*	ThermoFisher: Custom TaqMan® SNP Genotyping Assays	PTM08	
Commercial assay, kit	Taqman probe: WBVar00209467	ThermoFisher: Custom TaqMan® SNP Genotyping Assays	PTM11	
Commercial assay, kit	TruSeq Stranded mRNA kit	Illumina	Cat. No.: 20020595	
Commercial assay, kit	Zymo DNA isolation kit	Zymo	Cat. No.: D4071	
Commercial assay, kit	Zymo DNA cleanup kit	Zymo	Cat. No.: D4064	
Commercial assay, kit	ddPCR Supermix for Probes	BIORAD	Cat. No.: 1863010	
Commercial assay, kit	Droplet Generation Oils	BIORAD	Cat. No.: 1863005	
Commercial assay, kit	ddPCR Droplet Reader Oil	BIORAD	Cat. No.: 1863004	
Commercial assay, kit	VECTASHIELD antifade Mounting Medium with DAPI	VECTOR	Cat. No.: H-1200	

Table 1 – continued

Reagent type or resource	Designation	Source of reference	Identifiers	Additional information
Software, Algorithm	edgeR	PMID: 19910308	RRID:SCR_012802	Opensource: https://bioconductor.org/packages/release/bioc/html/edgeR.html
Software, Algorithm	SARTools	PMID: 27280887	RRID:SCR_016533	Opensource: https://github.com/PF2-pasteur-fr/SARTools
Software, Algorithm	MATLAB	MathWorks	RRID:SCR_001622	
Software, Algorithm	Rstudio	Rstudio	RRID:SCR_000432	https://www.rstudio.com/
Software, Algorithm	JMP12	SAS JMP	RRID:SCR_014242	
Software, Algorithm	Image J	NIH	RRID:SCR_003070	Opensource: https://imagej.nih.gov/ij/
Software, Algorithm	MetaMorph	Molecular Devices	RRID:SCR_002368	
Software, Algorithm	Custom TaqMan® Assay Design Tool	ThermoFisher		https://www.thermofisher.com/order/custom-genomic-products/tools/genotyping/

Strains

The following strains were used in this study:

Wild strains: N2; CB4856; DL238; JU258; JU775; MY16; MY23; CX11314; LKC34; ED3017; JT11398; EG4725. The N2 strain originated from the Bargmann lab (The Rockefeller University). The remaining eleven wild strains came from the *Caenorhabditis elegans* Natural Diversity Resource [118].

Barcoded strains: PTM229 *dpy-10 (kah82)II*; PTM288 *dpy-10 (kah83)II kyIR1(V, CB4856>N2) qgIR1(X, CB4856>N2)*; PTM289 *dpy-10 (kah84)II kyIR1(V, CB4856>N2) qgIR1(X, CB4856>N2)*; The barcoded strains were generated using previously published reagents for modifying the *dpy-10* gene [106]. Two modified repair oligos with the following sequence were used to edit silent mutations into the 90th (Thr) or 92nd amino acid (Arg):

dpy-10 90th silent mutation:

5'-
CACTTGAACTTCAATACGGCAAGATGAGAATGACTGGAAACCGTACTGCTCG
TGGTGCCTATGGTAGCGGAGCTTCACATGGCTTCAGACCAACAGCCTAT-3'

dpy-10 92nd silent mutation:

5'-
CACTTGAACTTCAATACGGCAAGATGAGAATGACTGGAAACCGTACCGCTCG
CGGTGCCTATGGTAGCGGAGCTTCACATGGCTTCAGACCAACAGCCTAT-3'

The microinjection mix was: 50ng/uL P*eft3::Cas9*, 25ng/uL *dpy-10* sgRNA, 500nM *dpy-10(cn64)* repair oligo, and one of the 500nM *dpy-10(90/92)* repair oligo. This mix was injected into N2 or CX12311 and so-called "jackpot broods" were identified by the presence of a large number of F1 animals with the roller phenotype. From these plates, wildtype animals were singled and genotyped using Sanger-sequencing. *kah82* and *kah83* contain the 90th Thr silent mutation (ACC -> ACT). *kah84* contains the 92nd Arg silent mutation (CGT -> CGC).

Near isogenic lines: CX12311 *kyIR1(V, CB4856>N2) qgIR1(X, CB4856>N2)*; QG1 *qgIR1(X, CB4856>N2)*; CX10774 *kyIR1(V, CB4856>N2)*; CX11400 *kyIR9(X, N2>CB4856)*. These strains were originally described in previous studies [19,71,107,108].

npr-1 loss of function: CX4148 *npr-1(ky13)X*; DA609 *npr-1(ad609)X*; These strains were previously described [75]

URX, AQR, PQR genetic ablation strains: qaIs2241*[Pgcy-35::GFP Pgcy-36::egl-1 lin15+]* is an integrated transgene that genetically ablates URX, AQR, and PQR neurons [5]. This transgene was crossed into a number of introgressed regions using standard genetic techniques. CX7102 *qaIs2241X*; PTM400 *qaIs2241X*; PTM401 *qgIR1(X, CB4856>N2) qaIs2241X*; PTM402 *kyIR1 (V, CB4856>N2) qaIs2241X* ; PTM403 *kyIR1(V, CB4856>N2) qgIR1(X, CB4856>N2) qaIs2241X* ;

daf-22 strains: *daf-22(kah8)II* is a *daf-22* gene disruption made by CRISPR/Cas9 genome editing [105]. The deletion of *daf-22* spans the 6th Pro to 219th Glu (The deleted sequence is: 5'-caaaggtatacatcgttggagtcggtatgacaaagttttgtaagccggga…ggatcaggtgatcaatgcccgtaagatctacgactttatgggtctcctcg-3'). This transgene was crossed into a number of introgressed regions using standard genetic techniques. PTM95 *daf-22(kah8)II kyIR1(V, CB4856>N2) qgIR1(X, CB4856>N2)*; PTM404 *daf-22(kah8)II dpy10(kah83)II*; PTM405 *daf-22(kah8)II*; PTM408 *daf-22(kah8)II kyIR1(V, CB4856>N2) qgIR1(X, CB4856>N2)*.

Growth conditions

Animals were grown following standard conditions. With exceptions listed below, animals were cultivated on modified nematode growth medium (NGM) plates containing 2% agar seeded with 200 μl of an overnight culture of the *E. coli* strain OP50 in an incubator set at 20°C. Strains were grown for at least three generations without starvation before any assays were conducted. For assays manipulating the environmental O_2 levels, animals were grown inside a BioSpherix C474 chamber using a BioSpherix C21 single chamber controller to control ambient O_2 levels. For these assays, animals were not grown in temperature incubators, and the room temperature was typically kept ~21°C. For competition experiments on non-burrowing plates, 1.25% agarose and 0.75% agar replaced the agar concentrations of normal growth plates. To create uniform lawns, liquid cultures of OP50 bacteria were poured onto plates to cover the entire surface area of the plate and then poured off.

UV treatment: 9cm NGM plates were seeded with 300uL of an overnight culture of the *E. coli* strain OP50 and placed at room temperature for two days. Then plates were placed in Stratagene UV Stratalinker 2400 with 254-nm radiation. The lids were removed and the plates were irradiated at 9,999mJ/cm^2. The efficacy of the killing was measured as described previously [115].

Pairwise fitness measurements

Competition experiments were performed as previously [105]. Briefly, Ten L4 stage animals from each strain were picked onto 9cm NGM plates seeded with 300uL of an overnight *E. coli* OP50 culture and incubated at room temperature for three days. After five days, animals were transferred to an identically-prepared NGM plate and then subsequently

transferred every four days for five to seven generations. For transfers, animals were washed off from the test plates using M9 buffer and collected into 1.5mL centrifuge tube. The animals were mixed by inversion and allowed to stand for approximately one minute to settle adult animals. 50uL of the supernatant containing ~1000-2000 L1-L2 animals were seeded on next plates. The remaining animals were concentrated and placed in a -80^{o}C freezer for future genomic DNA isolation. Genomic DNA was collected from every odd generation using a Zymo DNA isolation kit (D4071).

To quantify the relative proportion of each strain, we used a digital PCR based approach using a custom TaqMan probe (Applied Biosciences). Genomic DNA was digested with EcoRI for 30 min at 37 ^{o}C. The digested products were purified using a Zymo DNA cleanup kit (D4064) and diluted to ~1ng/ L for the following Taqman assay. Four TaqMan probes were designed using ABI custom software that targeted the *dpy-10 (kah82), dpy-10 (kah84), npr-1(g320),* or SNP WBVar00209467 in *glb-5*. These probes were validated using defined concentrations of DNA from animals containing each allele. The Taqman digital PCR assays were performed using a Biorad QX200 digital PCR machine with standard probe absolute quantification protocol. The relative allele proportion was calculated for each DNA sample using count number of the droplet with fluorescence signal (equation 1). To calculate the relative fitness of the two strains using three to four measurements of relative fitness, we used linear regression to fit this data to a one-locus generic selection model (equation 2 and 3), assuming one generation per transfer.

$$P(A)_t = \frac{No.Allele\ A}{No.Allele\ A + No.Allele\ a} \qquad (1)$$

$$P(A)_t = \frac{P(A)_0 {W_{AA}}^t}{P(A)_0 {W_{AA}}^t + (1 - P(A)_0){W_{aa}}^t} \qquad (2)$$

$$log(\frac{\frac{P(A)_0}{P(A)_t} - P(A)_0}{1 - P(A)_0}) = (log(\frac{W_{aa}}{W_{AA}}))t \qquad (3)$$

Aerotaxis assays

To measure bordering rates, two-week old NGM plates were removed from a 4°C cold room, seeded with 200 µL of *E. coli* OP50 and incubated for two days at room temperature. 150 adult animals were picked onto these assay plates and placed in either a 20°C incubator or a BioSpherix chamber for 3 hours. Bordering behavior was quantified using a dissecting microscope by identifying animals whose whole body resided within 1mm of the border of the bacteria lawn.

Transcriptome analysis

N2 and CX12311 L4 hermaphrodites were picked to fresh agar plates. Their adult progeny were synchronized using alkaline-bleach to isolate eggs. These eggs were washed three times using M9 buffer and placed on a tube roller overnight to allow eggs to hatch. About 400 L1 animals were placed on NGM agar plates seeded with non-uniform lawns of *E. coli* OP50 and incubated in a BioSpherix chamber set at 10% O_2 or 21% O_2 levels for 48 hours. The ~L4 stage animals were washed off and used for standard Trizol RNA isolation. Replicates were performed on different days. The RNA libraries for next-generation sequencing were prepared using an Illumina TruSeq Stranded mRNA kit (20020595) following its standard protocol. These libraries were sequenced using an Illumina NextSeq 500 platform. Reads were aligned using HISAT2 using default parameters for pair-end sequencing. Transcript abundance was calculated using HTseq and then used as inputs for the SARTools [146]. Within this R package, edgeR is used for normalization and differential

analysis. N2 cultured at 21% O_2 is treated as wild type [147]. The genes showing significantly different expression ($\log_2$(fold) > 1 or $\log_2$(fold) < -1, FDR adjusted p-value < 0.01) were selected to perform Hierarchical Cluster analysis, and Principal Component analysis. Sequencing reads were uploaded to the SRA under PRJNA437304.

Food consumption assays

Liquid food consumption: The 96 well-plates were prepared by pipetting 150 uL S media containing *E. coli* OP50 with density OD600 of 1.0 (CFU ~ $0.8x10^9$/mL), 500uM FUDR and 1 x Antibiotic-Antimycotic (ThermoFisher 15240062). 20 synchronized animals (L4 stage or young adult) were put into each well, pipetting to mix each well completely before using BioTek Synergy H4 multimode plate reader to record OD600 optical density every 24hrs from Day 1 to Day 5.

Plate food consumption: The 24 well-plates were prepared by pipetting 0.75mL NGM agar contain 25μM FUDR and 1x Antibiotic-Antimycotic (ThermoFisher 15240062) to each well. The freshly-prepared plates were placed in fume hood and dried with air flow for 1.5 hours. 20μL of freshly-cultured OD600 of 4.0 (CFU ~ 3.2×10^9/mL) *E. coli* OP50-GFP(pFPV25.1) were seeded in the center of each well. Animals were synchronized using alkaline-bleach. The eggs were washed by M9 buffer for three times and rotating on tube roller overnight to allow eggs to hatch. About 200 L1 animals were placed on NGM agar plates seeded with *E. coli* OP50 and cultivate at 20° C or BioSpherix chamber at 21° C for 50 hours. Ten animals (Late L4 stage or young adult) were transferred to each well of the first 5 columns of the food consumption assay 24 well-plates. The remaining 4 wells were

used as control wells to measure the GFP signal degradation. After place animals on the food consumption assay plates, the fluorescence signal of OP50-GFP from each well was quantified by area scanning protocol using BioTek Synergy H4 multimode plate reader at 6mm height as the starting time point. The 24-well plates were then incubated in incubator or BioSpherix chamber for 18 hours and the fluorescence signal were quantified again as the ending time point. The fluorescence signal at end time point from each well was normalized using the fluorescence signal degradation amount of control wells. The normalization was performed using the equation as below:

$$Fluorescence\_Control(0hr) = \beta \cdot Fluorescence\_Control(18hrs) \quad (1)$$

All of the signals from control wells were used to do linear regression and estimate coefficient β. The estimated amounts of bacteria at 18hrs for each test is:

$$Fluorescence(18hrs\_normalized) = \beta \cdot Fluorescence(18hrs) \quad (2)$$

The food consumption for each well was calculated by:

$$Food\ consumption\ amount = Fluorescence(0hr) - Fluorescence(18hrs\_normalized). \quad (3)$$

Pharyngeal pumping and size assays

Animals were synchronized using alkaline-bleach. The eggs were washed by M9 buffer for three times and rotating on tube roller overnight to allow eggs to hatch. About 200 L1 animals were placed on NGM agar plates seeded with *E. coli* OP50 and cultivated at 20^{o}C for 72 hours. In the pharyngeal pumping rates assays, the pharynges of ten young adult animals (72 hours after place L1 on NGM agar plate) were observed for 30 seconds each

in three separate trails. To measure the pharyngeal size, young adult animals were placed onto agar pad and immobilized by 25mM NaN_3. For each strain, pharyngeal sizes of 30 animals from three different plates were imaged under 40x objective lens using z-stack DIC microscope. The diameter of pharyngeal metacorpus, diameter of terminal bulb diameter, procorpus length, and isthmus length were measured using ImageJ software.

Reproductive timing and growth assays

To measure reproductive timing, animals were synchronized by picking ten adult animals onto an NGM plate, allowing them to lay eggs for two hours, and then removing the adult animals from the plate. These offspring were then monitored using a 12x dissecting microscope at indicated time points to count the number of animals with oocytes and fertilized eggs in their uterus. A subset of these animals was washed off at indicated time points and fixed in 95% ethanol. The nuclei were stained with 1.5 g/mL DAPI solution in Vectashield antifade mounting medium (VECTOR H-1200) for 10 minutes in the dark before visualization. Each spermatheca was imaged by z-stack fluorescence microscopy using a 100x lens to determine whether spermatogenesis had started or to count the number of sperm produced by the hermaphrodite.

Reproductive rate and body size measurements were measured as described previously [105].

Modifications for pheromone assays

For crude pheromone assays, crude pheromone was prepared as described previously [148]. The crude pheromone was resuspended in ethanol and stored in -20 ^{o}C. A dauer formation assay was performed to test the efficacy of crude pheromone. A 1/333 (v/v) crude

pheromone level could induce a high >80% rate of dauer formation in N2 animals grown on 20uL of heat killed *E. coli* OP50 bacteria (5mg/mL).

For the feeding assay and reproductive timing assays, the animals were grown on NGM plates for three generations on plates containing the indicated concentrations of crude pheromone (or ethanol control). + indicates a 1/10000 (v/v) crude pheromone and ++ indicated a 1/2000 (v/v) ratio of crude pheromones. The plates were then dried in biosafety cabinet for 1.5 hours, seeded with 200 µL of overnight culture of the *E. coli* strain OP50, incubated overnight, and used immediately for experiments.

Exploration assay

35 mm Petri dishes evenly seeded with OP50 *E. coli* Bacteria for 24h before the start of assay. Individual L4 hermaphrodites were placed in the center of the plate and cultivated in BioSpherix chamber in 10% O_2 or 21% O_2 level at 21° C for 3 hours. The plates were placed on a grid that has 105 squares which cover the whole plate. The number of full or partial squares that contained animal's tracks was quantified and the exploration fraction was calculated (equation 1).

$$Exploration\ fraction = \frac{No.\ grids\ contained\ tracks}{105} \qquad \textbf{(1)}$$

Statistics

All replicates were biological replicates using animals grown independently for multiple generations. The number of biological replicates were chosen using power analysis based upon the standard deviation from previous assays. To assess statistical significance, we performed one-way ANOVA tests followed by Tukey's honest significant difference test

to correct for multiple comparisons or the Wilcoxon-Mann-Whitney nonparametric test for pairwise comparisons. The Friedman test was used to compare the reproductive timing assays. The exact test used is listed in the legend for each panel.

Video Files

These files show a single generation (3 days) of growth of the N2 or CX12311 grown strain in the presence of 21% or 10% environmental O_2.

CHAPTER 3. A BENEFICIAL GENOMIC REARRANGEMENT CREATES MULTIPLE VERSIONS OF CALCIPRESSIN IN *C. ELEGANS*

Yuehui Zhao[1], Jason Wan[2], Daehan Lee[3], Erik C. Andersen[3], Hang Lu[2], Fred O. Vannberg[1,5], and Patrick T. McGrath[1,4,5]

[1]Department of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA

[2]School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA

[3]Department of Molecular Biosciences, Northwestern University, Evanston, IL, 60208, USA

[4]Department of Physics, Georgia Institute of Technology, Atlanta, GA 30332, USA

[5]Institute of Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, GA 30332, USA

3.1 Abstract

Gene duplication is a major source of genetic novelty and evolutionary adaptation, providing a molecular substrate that can generate biological complexity and diversity[149,150]. Despite an abundance of genomic evidence from extant organisms to the importance of gene duplication, consensus about how they arise and fix in a species is lacking[151]. In the process of studying the adaptation of laboratory strains of *C. elegans* to new food sources, we identified a recombinant inbred line (RIL) with higher relative fitness and hyperactive exploration behavior compared to either parental strain. Using bulked segregant analysis and short read resequencing, we identified a de novo beneficial, complex rearrangement of the *rcan-1* gene, which we resolved into five new unique tandem inversion/duplications using Oxford Nanopore long-read sequencing. *rcan-1* encodes an ortholog to the human *RCAN1/DSCR1* gene, which has been implicated as a causal gene for Down syndrome[152]. The genomic rearrangement in rcan-1 results in two complete and two truncated versions of the rcan-1 coding region, with a variety of modified promoter and 3' regions. Our results demonstrate that adaptation can occur through unexpectedly complex genetic changes that can provide the substrate for future evolutionary change.

3.2 Introduction and Results

3.2.1 An outlier RIL shows higher exploration behavior and fitness

Evolution is relentless, even in strains of model organisms used to study fundamental biological processes in laboratory settings. Despite supplying food and providing environments free of predators, animals adapt to their new and unnatural conditions [81,92-97,105,153]. One example is the standard laboratory strain of *C. elegans*, N2, which was isolated in 1951 from mushroom compost (**Figure 3.2.1a**). This strain spent almost two decades growing in the laboratory before methods of cryopreservation were developed [70]. Mutations that arose and fixed in N2 after 1958 can be identified using a second lineage that separated from the N2 population at that time, called LSJ2 [71]. Approximately 300 mutations distinguish these two strains. Previous studies have identified beneficial mutations that arose and fixed in the N2 lineage in the *glb-5*, *nath-10*, and *npr-1* genes [81,153].

To determine if any of other fixed mutations in the N2 lineage could affect fitness in standard laboratory conditions, we used a previously described panel of 89 recombinant inbred lines (RILs) between the CX12311 and LSJ2 strains. CX12311 is a near isogenic line that carries ancestral *npr-1* and *glb-5* alleles from the CB4856 Hawaiian wild strain introgressed into an N2 background (**Figure 3.2.1b-c** - henceforth referred to as N2*) [71]. Using N2* as a parental strain eliminates the fitness effect of the derived alleles of N2 *npr-1* and *glb-5* [153]. For each of the RIL strains, we performed pairwise competition experiments against the N2* strain to measure its relative fitness (**Figure 3.2.2**).

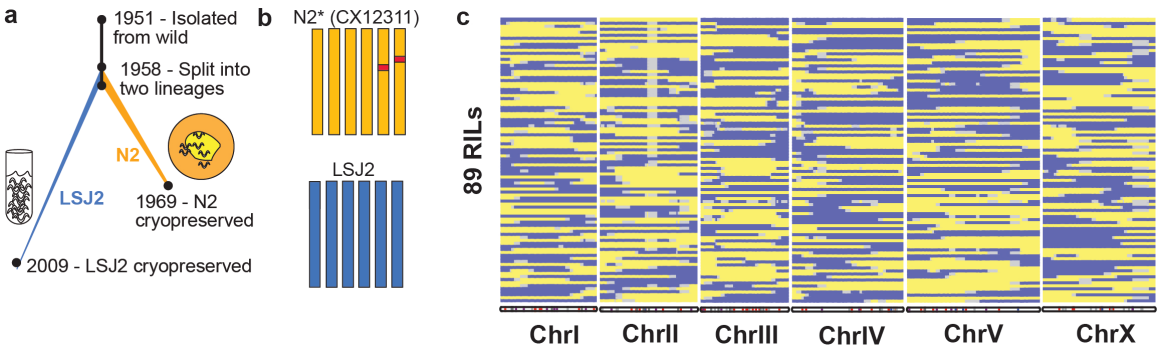

Figure 3.2.1. Genotype information of N2* ax LSJ2 recombinant inbred lines. **a**. History of laboratory domestication of N2 and LSJ2. **b**. Genotype information of the parental strains N2* and LSJ2. N2* (CX12311) carries introgressed CB4856 *glb-5* and *npr-1*. Orange represents N2 genetic component; blue represents LSJ2 genetic component; red represents CB4856 genetic component. **c**. Genotype information of N2* x LSJ2 recombinant inbred lines.

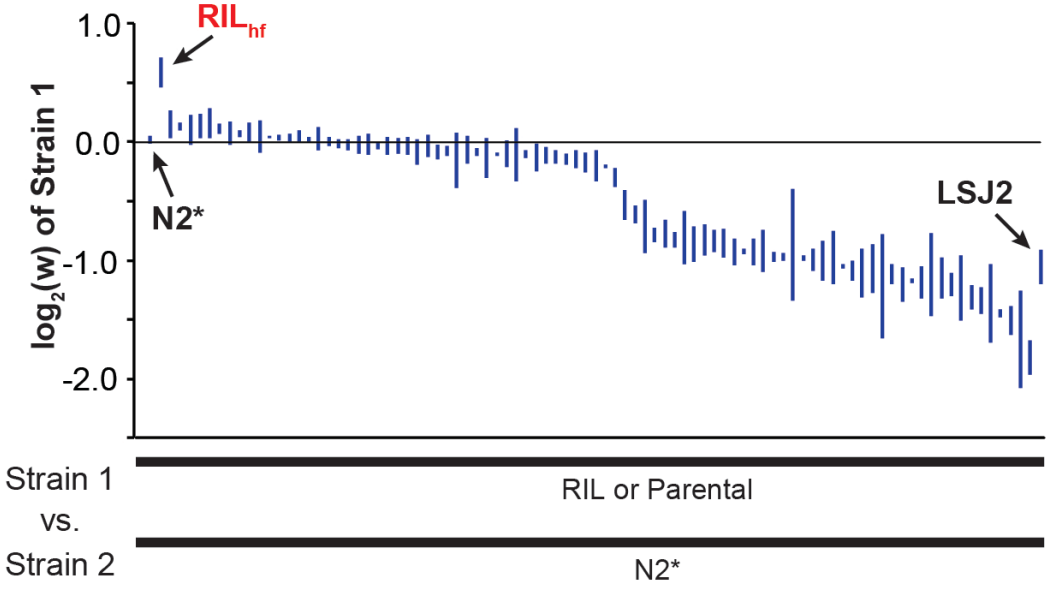

Figure 3.2.2. Relative fitness of N2* x LSJ2 RILs compare to N2*. Pairwise competition experiment of each 89 N2* and LSJ2 RILs compete with N2*. Each blue bar represents

the $\log_2$(w) value range of 2 or 3 replicates of each RIL. The RIL_{hf} indicates the high relative fitness RIL strain.

Interestingly, we found that 1 of the 89 RILs, CX12348 (henceforth called RIL_{hf} – hf for high fitness), had significantly higher fitness than either of the parental strains, which we validated in an independent experiment (**Figure 3.2.3a**). RIL_{hf} contained an approximately equal mixture of DNA from both the N2* and LSJ2 parental strains, with N2* DNA on the left arm of Chromosome I, the entire chromosomes II, III, and V, and portions of the X chromosome (**Figure 3.2.3b**). We decided to focus on this unusual RIL_{hf} to determine the genetic basis of its higher fitness than either the N2* or LSJ2 parental strains, reasoning that a *de novo* mutation that occurred during construction of the RILs could be responsible for higher fitness.

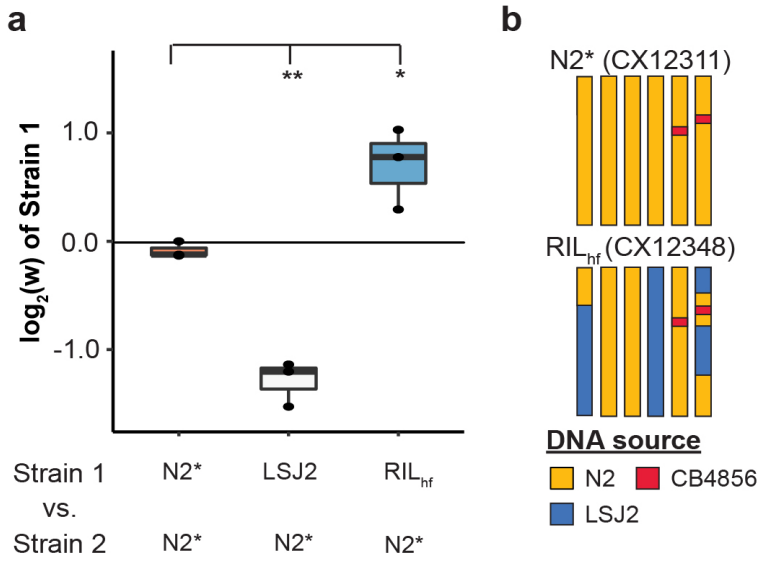

Figure 3.2.3. RIL strain CX12348 shows higher relative fitness.

a. RIL strain CX12348 shows higher relative fitness. *$*p<0.05$*, *$**p<0.01$* by ANOVA with Tukey's Honest Significant Difference test. **b.** Genotype information of CX12348.

Wild strains of *C. elegans*, as well as the N2* and LSJ2 parental strains, feed in groups on the borders of bacterial lawns, a foraging strategy known as social behavior [75]. While growing the RIL_{hf} strain on standard plates, we noticed that animals had a stronger propensity to explore the centers and regions outside of the bacterial lawns, resulting in an increased number of worms and tracks in the center and outside of the lawn (**Figure 3.2.4a**). In order to quantify this behavioral difference, we modified a previously described exploration assay [154] to measure long-term (16h) exploration behavior in the presence of circular lawns. The RIL_{hf} strain explored a substantially larger fraction of the bacterial lawn than either of the parental strains (**Figure 3.2.4b**). The exploration assay is an indirect measure of the relative time *C. elegans* spends in roaming and dwelling states in the presence of O_2 gradients (and other chemical gradients) created by the circular bacterial lawns.

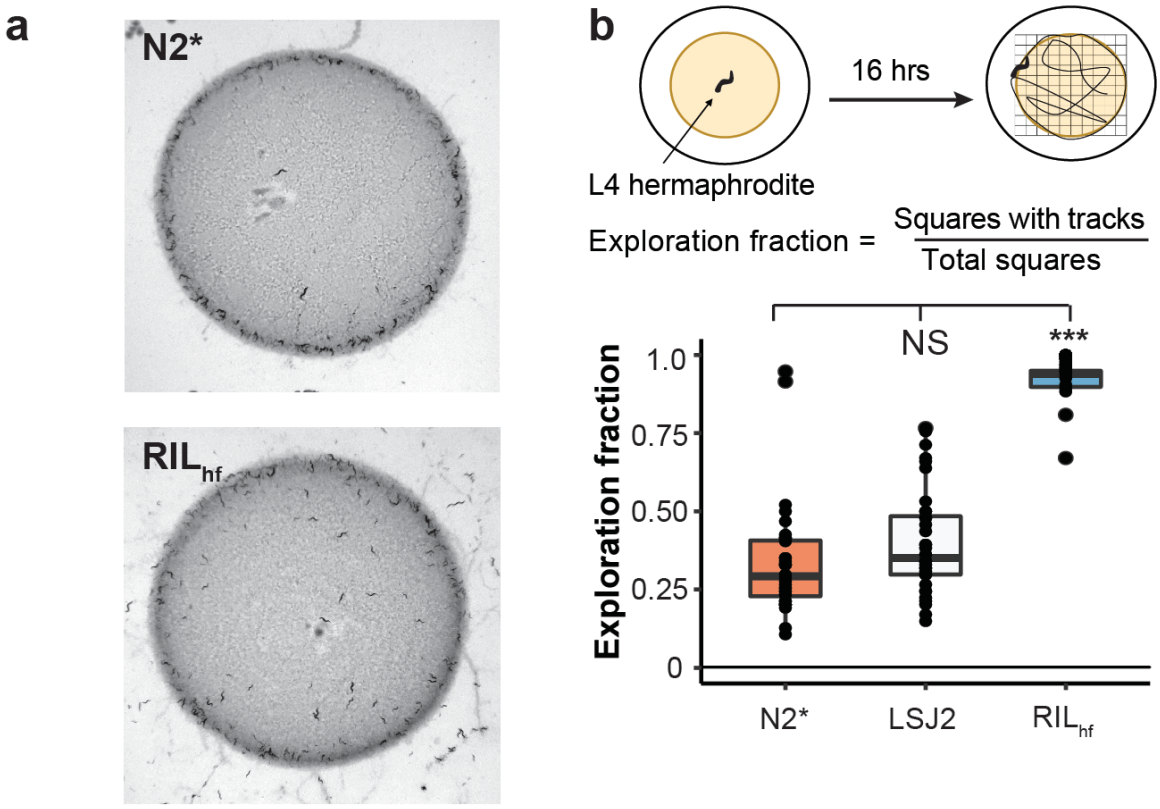

Figure 3.2.4. RIL$_{hf}$ strain shows active exploration behavior. a. Pictures of N2* and RIL$_{hf}$ grow on agar plates. **b.** RIL$_{hf}$ shows higher exploration behavioral fraction. Top illustrative shows the scheme of exploration behavioral assay. NS, not significant. ****p<0.001* by ANOVA with Tukey's Honest Significant Difference test.

3.2.2 Exploration behavior differences map to a region on Chromosome III

This change in foraging behavior is potentially an adaptive strategy for RIL$_{hf}$ animals to increase their evolutionary fitness in the laboratory or as a pleiotropic effect of the underlying genetic basis of this fitness gain. Since this foraging trait is easier to genetically map than relative fitness, we first focused on this phenotype. We created two panels of 48

recombinant inbred lines between RIL_{hf} and either the N2* or LSJ2 parental strains and measured their foraging behavior (**Figure 3.2.5a, b**). We grouped these strains into low or high exploration groups for each RIL panel and performed pooled genomic sequencing on the four groups that were created (**Figure 3.2.5c**). A large allelic imbalance between the high and low exploration groups was observed in the center of chromosome III in the RIL_{hf} x LSJ2 panels (**Figure 3.2.5d**). This result is expected if a *de novo* mutation occurred in RIL_{hf} in the center of chromosome III (which contains the N2 haplotype for the entire chromosome).

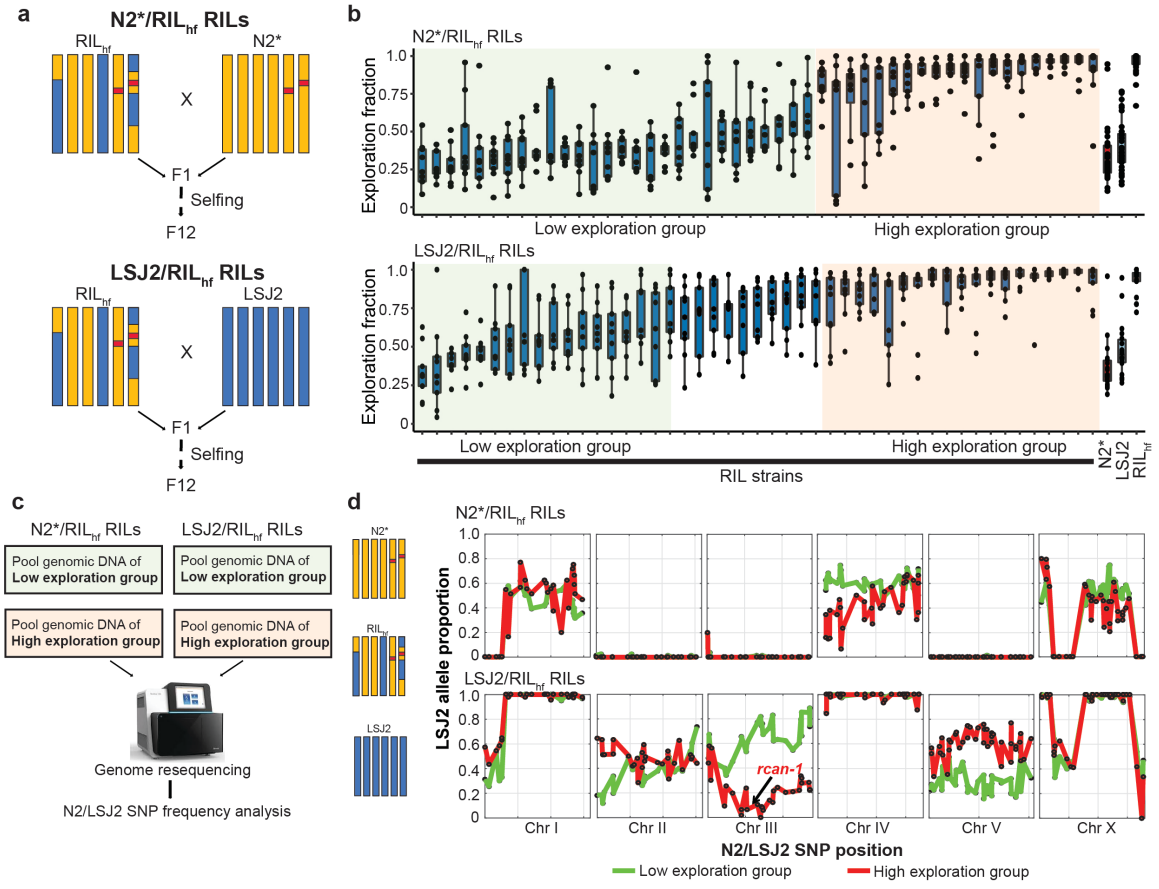

Figure 3.2.5. Gene mapping of exploration behavioral difference target to a region on Chromosome III. a. The scheme constructs N2*/ RIL_{hf} RILs and LSJ2/ RIL_{hf} RILs for

gene mapping. **b.** Exploration behavior of N2*/ RIL_{hf} RILs and LSJ2/ RIL_{hf} RILs. Each RIL were measured by 9 replicates on 3 different days. In either N2*/ RIL_{hf} RILs group or LSJ2/ RIL_{hf} RILs group, the RILs were assigned into high exploration group (red background) or low exploration group (green background). **c.** The scheme of exploration behavior bulk segregant analysis. d. Mapping the casual genetic loci using LSJ2 allele frequency in bulk segregant analysis. LSJ2 allele frequency was quantified at 200 N2*/LSJ2 genetic variants loci across the genome.

3.2.3 A complex rearrangement occurred at rcan-1 gene

To determine if the RIL_{hf} strain contains any *de novo* mutations in this region, we sequenced the RIL_{hf}, N2*, and LSJ2 strains using Illumina short read sequencing. While we did not identify any *de novo* SNVs or small indels on chromosome III in the RIL_{hf} strain, we did identify a large increase in coverage (2x – 8x) in the *rcan-1* gene, which is located in the center of chromosome III (**Figure 3.2.6a**). This coverage increase was detected in the high exploration groups of both RIL panels, consistent with this genetic change being causative for exploration behavior (**Figure 3.2.6a**).

What is the nature of the genetic change in the *rcan-1* region? While a simple gene duplication event would result in an increase in coverage, we observed a non-uniform change in coverage across the affected region. We also identified a large number of chimeric reads (reads which align to two unique locations with little overlap) that mapped

to five unique locations within the *rcan-1* locus. Together, these results suggest that the genetic change consisted of multiple inversion and/or duplication events. To resolve this change, we first attempted to amplify the entire affected region using PCR without success. As an alternative approach, we sequenced the RIL_{hf} strain using an Oxford Nanopore sequencing MinION, a long-read single molecule sequencing device with reported read lengths that could resolve the complex rearrangement. By identifying the reads that mapped to the *rcan-1* region, we identified a single X base long read that spanned the entire region (**Figure 3.2.6b**). By combining the Illumina short read and Nanopore long read data, we resolved the complex rearrangement into five unique tandem inversions interspaced within the *rcan-1* locus (**Figure 3.2.6c**).

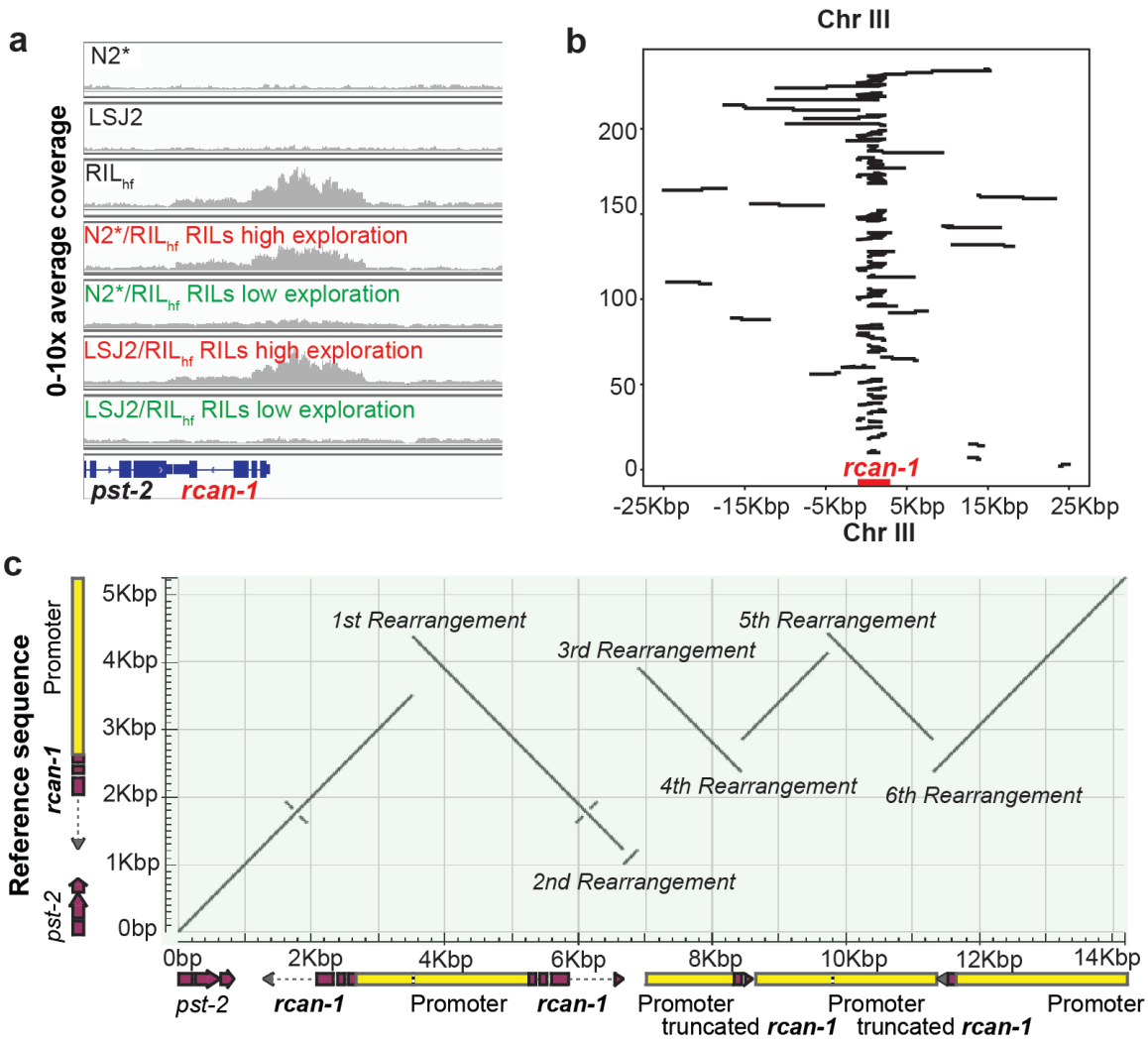

Figure 3.2.6. A complex genomic rearrangement occurred at *rcan-1* gene. a. Illumina sequencing shows sequencing coverage increases 2x – 8x at Chr III *rcan-1* gene region. **b.** The reads from long-read nanopore sequencing map to the *rcan-1* gene region. The zigzag shape of each read indicate the multiple gene inversion events. **c.** Pairwise alignment between wild-type *rcan-1* gene (y-axis) and *rcan-1* complex genomic rearrangement (x-axis).

3.2.4 The complex rearrangement is responsible for behavioral change and fitness increase

To determine if this rearrangement was responsible for the increases in exploration behavior and relative fitness of the RIL_{hf} strain, we created two near isogenic lines (NILs) by backcrossing the *rcan-1* rearrangement from the RIL_{hf} strain into the N2* background. As expected, both of these NILs explored a higher fraction of the bacterial lawn (**Figure 3.2.7a**). Pairwise competition experiments between the NILs and the N2* strain also demonstrated that this rearrangement is associated with the increases in the fitness of the RIL_{hf} strain (**Figure 3.2.7b**). The growth rates assay shows the body size difference between N2*, RIL_{hf} and NIL strains (**Figure 3.2.7c**).

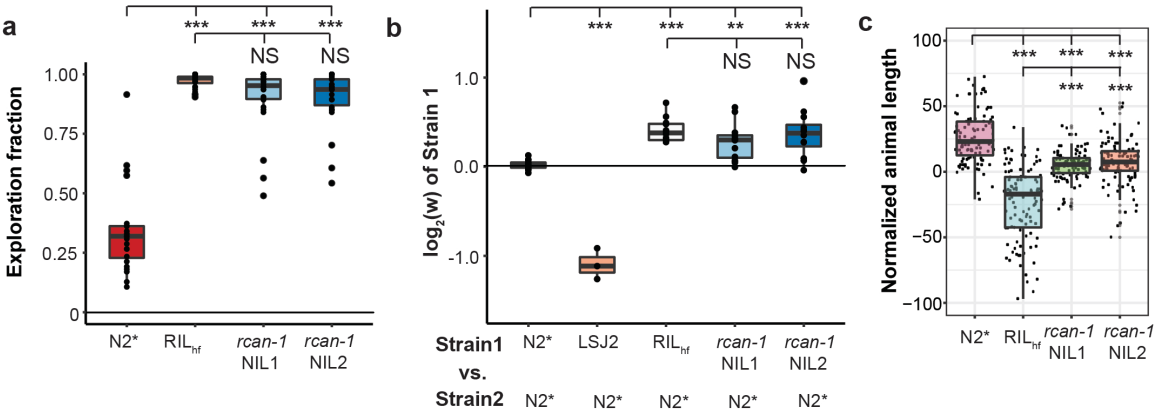

Figure 3.2.7. The complex genomic rearrangement is responsible for behavioral difference and fitness increase of RIL_{hf}. a. *rcan-1* NIL strains show high exploration behaviors. ****p<0.001* by ANOVA with Tukey's Honest Significant Difference test. **b.** *rcan-1* genomic rearrangement increase fitness benefit. NS, not significant.***p<0.01*, ****p<0.001* by ANOVA with Tukey's Honest Significant Difference test. **c.** High-

throughput growth rates assay. ***$p<0.001$ by ANOVA with Tukey's Honest Significant Difference test.

The *rcan-1* rearrangement is predicted to result in a number of changes to the *rcan-1* gene. First, it creates two full-length versions of the *rcan-1* coding region (**Figure 3.2.6**). However, the promoter region for each of these is modified by an inversion event in between the two coding regions. The first and second copy of *rcan-1* contain 857 and 1,725 bp of the normal promoter before the inversion event occurs, respectively. Given that the next gene is over 5kb upstream of the *rcan-1* gene in wildtype animals, enhancers and other regulatory regions of the promoter might be missing in the rearranged region, resulting in decreased, increased, or ectopic expression of *rcan-1*. Second, the second copy of the *rcan-1* gene also contains a small inversion in the 3' UTR region. This inversion could modify binding sites for small RNAs or other RNA-binding proteins that regulate the stability or translation of the mRNA product. Finally, two truncated copies of the *rcan-1* gene are also created, containing the first two exons of the gene. It is possible that these products are translated into proteins that could modify cellular function in RCAN-1 dependent or independent ways. *A priori*, it is difficult to predict which of these changes acting alone or in concert could result in the changes to foraging behavior and/or evolutionary fitness.

To gain insight into this, we transcriptionally profiled the RIL_{hf} strain and compared it to N2* animals. Interestingly, the gene with the largest change in expression was *rcan-1*,

suggesting that the rearrangement resulted in decreased expression of the *rcan-1* gene (**Figure 3.2.8 and table 2**).

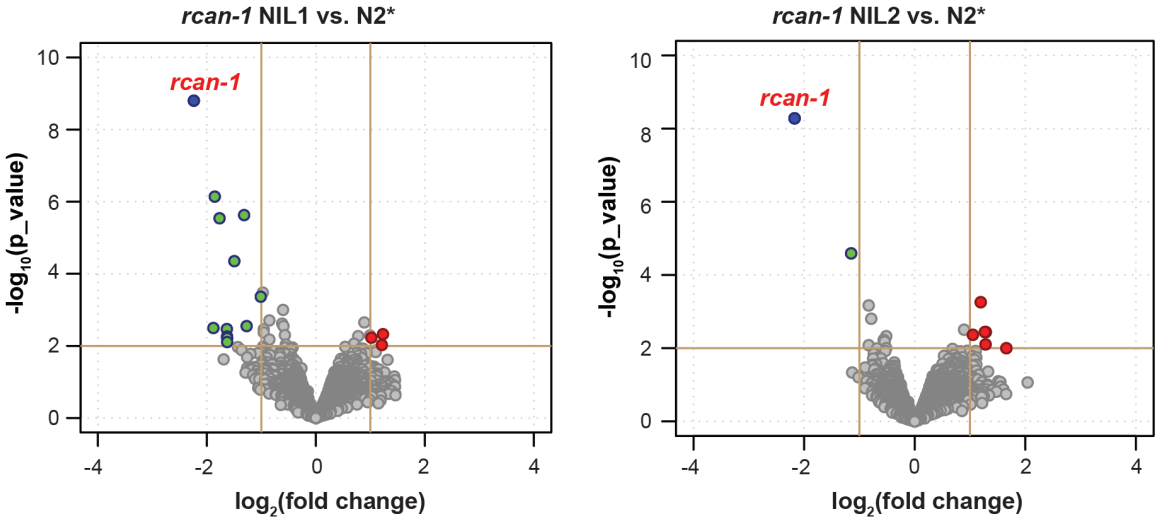

Figure 3.2.8. Transcriptome analysis of *rcan-1* NIL strains. Volcano plots shows *rcan-1* NILs differential gene expression pattern compare to N2*. Green color indicates the genes show $\log_2$(fold change) < 0.5 and $-\log_{10}$(p_value) > 2; Red color indicates the gene show $\log_2$(fold change) > 2 and $-\log_{10}$(p_value) > 2. The *rcan-1* gene marked as blue color show the largest gene expression decrease.

3.2.5 The residual expression of rcan-1 genes are responsible for the fitness benefits and exploration behavioural changes

Since the transcriptional profiling only reports whole body changes in total *rcan-1* expression, we created promoter::mCherry fusions to determine how the two modified promoters for the full-length *rcan-1* gene affected expression in the complex rearrangement. We cloned the entire promoter region between the two full-length versions of *rcan-1* in both directions to create a $P_{rcan\text{-}1\text{-}R1}$::mCherry construct (reporting expression

of the first full-length version of *rcan-1* in the complex rearrangement) and a $P_{rcan\text{-}1\text{-}R2}$::mCherry construct (reporting expression of the second full-length version of *rcan-1* in the complex rearrangement). As a control, we also cloned the wild-type 5085bp length promoter from N2 and fused it to GFP or mCherry ($P_{rcan\text{-}1\text{-}WT}$::GFP or $P_{rcan\text{-}1\text{-}WT}$::mCherry). We then simultaneously co-injected $P_{rcan\text{-}1\text{-}WT}$::GFP with $P_{rcan\text{-}1\text{-}WT}$::mCherry, $P_{rcan\text{-}1\text{-}RA\text{-}short}$::mCherry, or $P_{rcan\text{-}1\text{-}RA\text{-}long}$::mCherry. We used confocal microscopy to image whole-body expression from both green and red channels (**Figure 3.2.9a**). As expected from a previous publication, we observed wild-type expression of *rcan-1* in a variety of tissues, including motor, head, pharyngeal and tail neurons, pharyngeal cells, and hypodermal cells. We first measured how these promoters affected whole body expression by integrating the total GFP and mCherry signal from ~30 animals for each promoter construct (**Figure 3.2.9b**). Both of the promoters from the complex rearrangement drove less mCherry expression than the wild-type promoter to different extents, with the 1st promoter rearrangement more affected. While the 2nd promoter was able to drive expression at a higher level, it is important to note that these promoter fusion constructions lack the endogenous 3'UTR and transcriptional terminator sites. Combined with the whole body RNAseq data, we suggest that the *rcan-1* expression is largely reduced in the RIL_{hf} strain due to modifications in the promoter regions of the both full-length versions combined with the changes in the 3' end of the 2nd full-length copy of the *rcan-1* coding region.

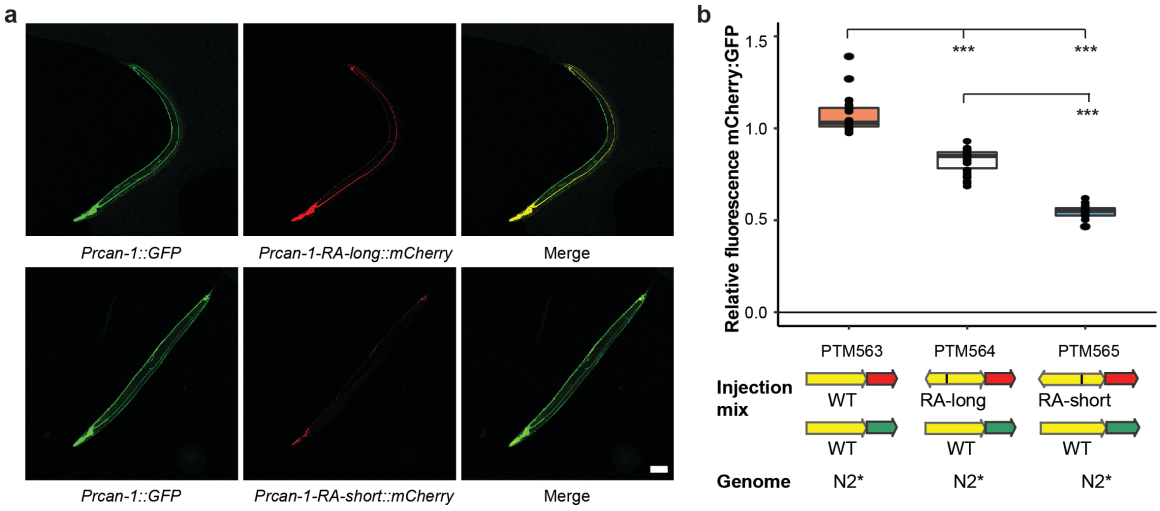

Figure 3.2.9. Expression analysis of two truncated *rcan-1* promoters. a. Confocal imaging suggests the expression difference between the long rearrangement promoter (RA-long 1725bp) and the short rearrangement promoter (RA-short 857bp). **b.** The relative expression pattern (mCherry/GFP) of each transgenic that inject the different construct showing in the figure. ****p<0.001* by ANOVA with Tukey's Honest Significant Difference test.

The above experiments suggest that the *rcan-1* rearrangement could be beneficial due to reduction or loss of transcription in specific tissues while retaining wild-type or 'enough' transcription in other tissues. Alternatively, the rearrangement could be beneficial because loss or *rcan-1* across the entire animal is beneficial. To test this, we used CRISPR-enabled genomic editing to delete *rcan-1* in the N2* strain. This knockout strain showed an intermediate phenotype between the N2* and the *rcan-1* NIL strains in the modified foraging behavior (**Figure 3.2.10a**). When we competed this strain against the two *rcan-1* NILs, the results demonstrated that the *rcan-1* rearrangement was substantially more fit

than the *rcan-1* deletion (**Figure 3.2.10b**). We conclude that the benefits of the *rcan-1* rearrangement are more complex than simply loss of *rcan-1* function.

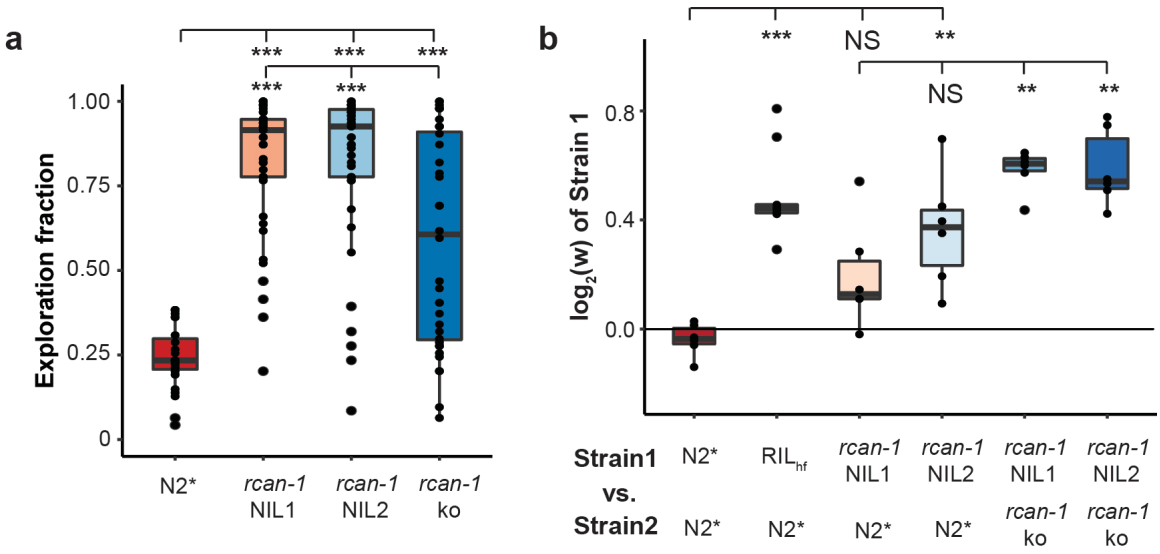

Figure 3.2.10. The complex rearrangements of *rcan-1* shows different effects from *rcan-1* knock out on behavioral changes and fitness benefits. a. The *rcan-1* gene knock out (*rcan-1 ko*) shows different effect from the complex rearrangements of *rcan-1*. ****p<0.001* by ANOVA with Tukey's Honest Significant Difference test. **b.** The complex rearrangements of *rcan-1* lead to higher fitness than *rcan-1ko.* NS, not significant. ***p<0.01*, ****p<0.001* by ANOVA with Tukey's Honest Significant Difference test.

3.3 Discussion

Our results demonstrate that evolution can occur rapidly and through unexpectedly complex genetic changes. The rearrangement of *rcan-1* occurred in a 12-generation pedigree used to construct 96 RIL lines. We propose that this rearrangement occurred as a

single event, potentially as a result of an error in Okazaki fragment ligation during DNA replication. *rcan-1* encodes an ortholog to the human *RCAN1* gene, which encodes a calcipressin family protein that inhibits the calcineurin A protein phosphatase. In mammals, the exact genetic dosage of *RCAN1* is important; it has been proposed to be a key contributor to Down Syndrome phenotypes in patients with trisomy 21 and chronic overexpression of *RCAN1* in mice promotes a number of phenotypes related to Alzheimer's disease. In *C. elegans*, less is known about the function of *rcan-1*. It has been shown to be required for memory of temperature exposure through a *tax-6*/calcineurin-family and *crh-1*/CREB-dependent pathway. Thermotaxis, however, is not predicted to be important for laboratory fitness, and it is possible that the *rcan-1* rearrangement regulates other unknown aspects of *C. elegans* biology that selection can act on. In mice, it was recently shown that *RCAN1* regulates fatty acid metabolism. If this is conserved in *C. elegans*, potentially the use of energy from food is diverted towards reproduction or other fitness-relevant traits.

The nature of the genetic change of *rcan-1* is perhaps the most interesting aspect of this report. While causative, genetic variants responsible for evolutionary change have been identified, these mutations tend to be SNVs, indels, or simple copy-number variants that modify protein function or expression. The rearrangement of *rcan-1* is more complicated and results in large-scale changes to promoter regions, 3'UTR regions, and the creation of truncated versions of the *rcan-1* gene. For short evolutionary timescales, this type of genetic variant could potentially access changes to cellular function that would be difficult for a single SNV, indel, or CNV to cause. For example, besides changing the exact tissues *rcan-1* gene is expressed in, it is also predicted to create an mRNA with a modified 3'

UTR, potentially decoupling translational regulation or modifying mRNA stability in a subset of tissues. From a macroevolutionary perspective, the generation of two full-length versions of a gene, each with different promoters and regulatory regions, provides a flexible substrate for potentially generating two paralogs with different functions. Further work identifying the cellular, molecular, and phenotypic changes responsible for increases in evolutionary fitness should be illuminating in this regard.

Genetic changes like this complex rearrangement are still difficult to identify using standard resequencing technologies and bioinformatic algorithms. As long-read sequencing technology improves, it will be interesting to determine the extent that similar alleles segregate within populations, and their relative importance in determining the unique forms and behaviors of species.

3.4 Materials and Methods

Nematode Growth Conditions

The animals were cultivated on standard nematode growth medium (NGM) plates containing 2% agar seeded with 200 µl of an overnight culture of the *E. coli* strain OP50 in an incubator set at 20°C. Strains were grown for at least three generations without starvation before any experiments were conducted.

Strains

The following strains were used in this study:

Near isogenic lines (NILs):

CX12311 (N2*) - *kyIR1(V, CB4856>N2), qgIR1(X, CB4856>N2)*;

PTM413 (*rcan-1* NIL 1) *kahIR16(III, CX12348>N2), kyIR1(V, CB4856>N2), qgIR1(X, CB4856>N2)*;

PTM414 (*rcan-1* NIL 2), *kahIR17(III, CX12348>N2), kyIR1(V, CB4856>N2), qgIR1(X, CB4856>N2)*;

Recombinant inbred lines (RILs):

The starting strains for generating these RILs are CX12311 and LSJ2. CX12311 is the near-isogenic line with CB4856 *glb-5* region at chromosome V and *npr-1* region at chromosome

X introgressed into N2 background. The details of construction and genotype information of the RILs are described in the previous study[71].

CX12311 – LSJ2 RILs: CX12312-19, CX12321-27, CX12346-52, CX12354-60, CX12362-66, CX12368-75, CX12381-88, CX12414-37, CX12495-99, CX12501-08, CX12510, CX12361

CX12311 males or LSJ2 males were crossed with high fitness CX12311-LSJ2 RIL CX12348 hermaphrodites. 95 F2 progeny (48 from CX12311 x CX12348; 47 from LSJ2 x CX12348) were cloned to individual plates and allowed to self-fertilized for 12 generations to create the inbred lines.

CX12311 - CX12348 (RIL_{hf}) RILs: PTM378-397, PTM421-434, PTM494-503.

LSJ2 - CX12348 (RIL_{hf}) RILs: PTM435-478.

CRISPR-generated knockout:

PTM505: *kyIR1 (V, CB4856>N2); qgIR1 (X, CB4856>N2) X; rcan-1(kah183) III dpy-10(kah83)II*

CRISPR-barcoded strain:

PTM288: *kyIR1(V, CB4856>N2), qgIR1(X, CB4856>N2) dpy-10(kah83)II*;

Extrachromosomal array strains:

PTM553 *kyIR1(V, CB4856>N2), qgIR1(X, CB4856>N2) Ex169 (pSM-Prcan-1::GFP 25ng/μL pTM8-Prcan-1::mCherry 25ng/μL).*

PTM554 *kyIR1(V, CB4856>N2), qgIR1(X, CB4856>N2) Ex170 (pSM-Prcan-1::GFP 25ng/μL pTM8-Prcan-1::mCherry 25ng/μL).*

PTM555 *kyIR1(V, CB4856>N2), qgIR1(X, CB4856>N2) Ex171 (pSM-Prcan-1::GFP 25ng/μL pTM8-Prcan-1::mCherry 25ng/μL).*

PTM559 *kyIR1(V, CB4856>N2), qgIR1(X, CB4856>N2) Ex175 (pSM-Prcan-1::GFP 25ng/μL pTM8-Prcan-RA-short::mCherry 25ng/μL).*

PTM560 *kyIR1(V, CB4856>N2), qgIR1(X, CB4856>N2) Ex176 (pSM-Prcan-1::GFP 25ng/μL pTM8-Prcan- RA-short::mCherry 25ng/μL).*

PTM561 *kyIR1(V, CB4856>N2), qgIR1(X, CB4856>N2) Ex177 (pSM-Prcan-1::GFP 25ng/μL pTM8-Prcan- RA-short::mCherry 25ng/μL).*

PTM556 *kyIR1(V, CB4856>N2), qgIR1(X, CB4856>N2) Ex172 (pSM-Prcan-1::GFP 25ng/μL pTM8-Prcan- RA-long::mCherry 25ng/μL).*

PTM557 *kyIR1(V, CB4856>N2), qgIR1(X, CB4856>N2) Ex173 (pSM-Prcan-1::GFP 25ng/μL pTM8-Prcan- RA-long::mCherry 25ng/μL).*

PTM558 *kyIR1(V, CB4856>N2), qgIR1(X, CB4856>N2) Ex174 (pSM-Prcan-1::GFP 25ng/μL pTM8-Prcan- RA-long::mCherry 25ng/μL).*

Molecular Cloning

To construct sgRNA vectors, the sgRNA sequence was inserted into a modified pDD162 vector (P*U6::sgRNA*) using NEB Q5 site directed mutagenesis kit (E0554).

To construct pSM-*Prcan-1*::GFP, *Prcan-1* is amplified from CX12311 genomic DNA using primers F: 5'-ctgGGCCGGCCtcggttcaaatacctcatgggaca-3' R: 5'-ttGGCGCGCCtttttgttgttaacttatagaaaaa atttcagcaacca-3' and cloned into pSM-GFP backbone with restriction enzyme sites 5'-FseI and 3'-AscI.

To construct pTM8-*Prcan-1*::mCherry, *Prcan-1* is amplified from CX12311 genomic DNA using primers F: 5'-tcggttcaaatacctcatgggaca-3' R: 5'- tttttgttgttaacttatagaa aaaatttcagcaacca-3'; PTM8 is amplified from PTM8-mCherry backbone using primers F: 5'- atttttttctataagttaacaacaaaaaacaagtttgtacaaaaaagcaggct-3' R: 5'-ccatgaggtatttgaaccgaatagcttggcgtaatcatggtcat-3'; The two fragments are assembled using HI-FI assembly (NEB E5520S).

To construct pTM8-*Prcan-1-RAlong*::mCherry and PTM8-*Prcan-1-RAshort*::mCherry, the rearrangement *Prcan-1* promoter region which driven two intact *rcan-1* expression is amplified from CX12348 genomic DNA. The long rearrangement promoter is amplified using primers F: 5'- tttttgttgttaacttatagaaaaaatttcagca-3' R: 5'- agcggacccaccttgtttc-3' and the short long rearrangement promoter is amplified using primers F: 5'-tttttgttgttaacttatagaaaaaatttcagca-3' R: 5'- gaaacgaaacaaggtgggtcc-3'. The PTM8-mCherry backbone is amplified using primers F: 5'-atttttttctataagttaacaa caaaaaatagcttggcgtaatcatggtcat-3' R: 5'- atttttttctataagttaacaacaaaaaacaagtttgtacaaaa aagcaggct-3'. The three fragments were assembled using HI-FI assembly and generated pTM8-*Prcan-1-RAlong*::mCherry and PTM8-*Prcan-1-RAshort*::mCherry. The plasmids were verified by enzyme digestion and Sanger sequencing.

Transgenic lines

Transgenic lines were made by microinjection using the standard protocol[155]. To make transgenic lines PTM553-561, *Prcan-1::GFP 25ng/μL* and *Prcan-1::mCherry 25ng/μL* were coinejected to PTM553-555; *Prcan-1::GFP 25ng/μL* and *Prcan- RA-short::mCherry 25ng/μL* were coinejected to PTM559-561; *Prcan-1::GFP 25ng/μL* and *Prcan-RA-long::mCherry* were coinejected to PTM556-558. The expression of *Prcan-1::GFP* in each transgenic line was used to calculate the relative mCherry expression ratio of *Prcan-1::mCherry*, *Prcan- RA-short::mCherry,* and *Prcan-RA-long::mCherry*.

Mutants

PTM505 strain: *rcan-1(kah183)III* is a gene disruption made by CRISPR/Cas9 genome editing. The deletion of *rcan-1* spans the 14th Ala to 205th Cys (The deleted sequence is: 5'-caatggatcattgatcagcactgtttccagtaaagatgatcttccaaatg…cgtaccatcagcaattgaaatgccacgaactccacgcccttcatctccat-3'). The microinjection mix was: 50ng/uL P*eft3::Cas9*, 10ng/uL *dpy-10* sgRNA, 25ng/uL *rcan-1* sgRNAs mixture target at N-terminus and C-terminus of gene body, 500nM *dpy-10(cn64)* repair oligo. This mix was injected into CX12311 and so-called "jackpot broods" were identified by the presence of a large number of F1 animals with the roller phenotype. From these plates, Dumpy or Roller animals were singled and genotyped using PCR. The strain was backcrossed to PTM288 strain.

The *rcan-1* sgRNA targeting sequences:

sgRNA1: 5'-tcaacgaaatccgttgccaa-TGG-3'; sgRNA2: 5'-agtgctgatcaatgatccat-TGG-3'; sgRNA3: 5'-cgtggcatttcaattgctga-TGG-3'; sgRNA4: 5'-tcacatggagatgaagggcg-TGG-3'.

Primer pair 1(Amplified from wildtype): F: 5'-ctcactttccttctgctttttttatgga-3'; R: 5'-gaaagcatttg ggaattggcact-3'. Primer pair 2(Amplified from wildtype): F: 5'-gctgacaccagcaatcattgttca-3'; R: 5'-ggcagaataagctctgcggat-3'. Primer pair 3(Amplified from gene disruption mutant): F: 5'-ctcactttc cttctgctttttttatgga-3'; R: 5'-ggcagaataagctctgcggat-3'. The line with the *rcan-1* gene disruption was backcrossed with PTM288 for 4 generations. The *rcan-1* disruption homozygote animals were genotyped with PCR and Sanger sequencing.

Competition experiment

Competition experiments were performed as previously. In brief, 9cm NGM plates seeded with 300uL of an overnight E. coli OP50 culture and incubated at room temperature for three days for bacteria growth. 10 L4 stage animals from each strain were picked onto the plates and cultured for 5 days. Then the animals were transferred to an identically-prepared NGM plate and then subsequently transferred every four days for five to seven generations. For each transfer, the animals were washed off from the plates using M9 buffer and collected into 1.5mL centrifuge tube. The animals were then mixed by inversion and allowed to stand for approximately for 1 minute to settle down adult animals. 50μL of the supernatant containing ~1000-2000 L1-L2 animals were seeded onto next plates. The remaining animals were concentrated and placed in a -80oC freezer for future genomic DNA isolation. Genomic DNA was collected from every odd generation using a Zymo

DNA isolation kit (D4071). To quantify the relative proportion of each strain, a Taqman assay was performed by digital PCR based approach with custom TaqMan probes (Applied Biosciences). Genomic DNA was digested with SacI or EcoRI for 30 min at 37 oC. The digested products were purified using a Zymo DNA cleanup kit (D4064) and diluted to ~1ng/μL- 2ng/μL for the following Taqman assay. Seven TaqMan probes were designed using ABI custom software that targeted the SNP WBVar00051876, SNP WBVar00601322, SNP WBVar00167214, SNP WBVar00601493, SNP WBVar00601538, *dpy-10* (kah83), and the SNP on *tbc-10* (reference genome WS268: ChrIII 985677 G/T). These probes were validated using the defined concentrations of DNA from strains containing each allele. The Taqman digital PCR assays were performed using a Biorad QX200 digital PCR machine with standard probe absolute quantification protocol. The relative fitness was calculated follow the equations described in the previous research[74]. The relative fitness value and Taqman assay information for each competition experiment are included in supplementary table 5.

Exploration behavioral assay

35 mm Petri dishes seeded with 150μL OP50 *E. coli* Bacteria for 24h before the start of assay. Individual L4 hermaphrodites were placed in the center of the plate and cultivated in 20 ^{o}C for 16 hours. The plates were placed on a grid that has 100 squares which cover the whole bacteria lawn. To calculate the exploration fraction, the number of full or partial squares that contained animal's tracks out of bacteria lawn was quantified. The number of

full or partial squares that contain the bacteria lawn was also counted (about 94-96 grids). The exploration fraction was calculated (equation 1).

$$Exploration\ fraction = \frac{No.\ grids\ contained\ tracks}{No.grids\ contained\ bacteria\ lawn} \qquad \textbf{(1)}$$

Bulk segregant analysis of exploration behavioral traits

The exploration behavioral assays were performed with all the 48 CX12311-CX12348 RILs (N2*/RIL_{hf} RILs) and 47 LSJ2-CX12348 RILs (LSJ2/RIL_{hf} RILs). In N2*/RIL_{hf} RILs, the 28 RILs show lower exploration behavior (exploration fraction median < 0.575) were assigned to low exploration group; the 20 RILs show higher exploration behavior (exploration fraction median ≥ 0.575) were assigned to high exploration group. In LSJ2/ RIL_{hf} RILs group, the 17 RILs show lower exploration behavior (exploration fraction median < 0.620) were assigned to low exploration group; the 20 RILs show higher exploration behavior (exploration fraction median ≥ 0.870) were assigned to high exploration group; the rest RILs were assigned to medium exploration group. The genomic DNA for each RIL was isolated. 100ng genomic DNA of each RIL from N2*/RIL_{hf} RILs group / LSJ2/ RIL_{hf} RILs group were pooled together into as high exploration group genomic DNA or low exploration group genomic DNA pool. The whole genome sequencing library of each genomic DNA were prepared and sequenced at 150bp paired-read with ~60x coverage. 200 N2*/LSJ2 SNPs were chosen for representing the frequency of LSJ2 allele (LSJ2 allele reads number/total reads number). The exploration fraction data for each RIL are included in supplementary table 5. The N2/LSJ2 allele proportion for selected markers in each RIL group are included in supplementary table 6.

Whole genome resequencing

Genomic DNA was isolated from each strain grown on 9cm NGM plates using Qiagen Gentra Puregene Kit (158667) following the supplementary protocol for nematodes. The genomic DNA were purified using Zymo Quick-DNA kit (D4068). The DNA libraries were prepared by Illumina Nextera DNA kit (FC-121-1030) and indexes (FC-121-1011). The prepared libraries were sequenced at 35bp or 150 bp paired-read by Illumina Nextseq 500. The quality of the fastq files was tested by fastqc. The reads were aligned to reference genome using BWA-aligner[156]. The output SAM files were converted to BAM files by Picard and then merged using SAMtools[157]. The merged BAM files were then deduplication and indexed using Picard[158]. The SNVs were called by Freebayes and annotated by SnpEff[159,160]. The clipped reads were used to identify the potential structural variants. The reads depth at *rcan-1* region were shown in software IGV[161]. The *de novo* point mutations and small insertion/deletion that detected in CX12348, PTM413, and PTM414 are included in supplementary table 7. The sequencing reads were uploaded to the SRA under SUB5262289.

***rcan-1* complex rearrangement analysis using Nanopore sequencing**

The genomic DNA of CX12348 was isolated from animals grown on 8 9cm NGM plates using Qiagen Gentra Puregene Kit (158667) following the supplementary protocol for nematodes. The genomic DNA were concentrated and purified using Zymo Quick-DNA kit (D4068). Size-selection to collect DNA fragments from 10kbp – 50kbp was carried out by Blue-pippin technique. The sequencing library was prepared using 1D ligation kit (SQK-LSK108) follow the standard protocol. In brief, DNA repair step was performed

using NEBNext FFPE Repair Mmix(M6630). After DNA repair, end preparation was performed and the adapter was ligated. ng prepared library was loaded in the Nanopore R9 flow cell in MinION sequencer. The standard 48 hours sequencing protocol was performed and about 5Gb sequencing data was generated. To resolve the structure of rcan-1 complex rearrangement, the fastq files were aligned to reference genome using BWA aligner. The reads cover the rcan-1 gene region were fetched using Pysam package in Python3. These reads were then mapped to rcan-1 gene region (50kbp) by blast and plotted by Matplotlib in Python3 to show the rearrangement events. The structure of the complex rearrangement was verified by illumina short reads sequencing and Sanger sequencing. The details of structure were represented by pairwise alignment using NCBI Blast tool. The sequencing reads were uploaded to the SRA under SUB5262289.

RNA-seq and transcriptome analysis

CX12311, PTM413, and PTM414 L4 hermaphrodites were picked to fresh agar plates. Their adult progeny were bleached using alkaline-bleach solution to isolate eggs for synchronization. The eggs were washed with M9 buffer for three times and placed on a tube roller overnight. About 400 hatched L1 animals were placed on NGM agar plates and incubated at 20°C for 48 hours. The ~L4 stage animals were washed off for standard RNA isolation using Trizol. Four replicates for each strain were performed on different days. The RNA libraries were prepared using an NEB Next Ultra II Directional RNA Library Prep Kit (E7760S) following its standard protocol. The libraries were sequenced by Illumina NextSeq 500. The reads were aligned by HISAT2 using default parameters for pair-end sequencing. Transcript abundance was calculated using HTseq and then used as inputs for the SARTools[146]. edgeR was used for normalization and differential analysis.

The analysis result was shown in volcano plot. CX12311 was treated as wild type[147]. The genes show significant differential expression in volcano plot are under thresholds $|\log_2(\text{fold})\} > 1$ and FDR adjusted p-value < 0.01. The sequencing reads were uploaded to the SRA under SUB5262289.

Imaging

The detailed steps of microdevice fabrication were previously reported[162]. Master molds were fabricated using UV photolithographic processes with a negative photoresist, SU8-2025 (Micochem). The microfluidic devices were fabricated in polydimethysiloxane (PDMS) (Dow Corning Sylgard 184) by soft lithography. Prior to micromolding, the 55 μm-thick patterned wafers were treated with trichloro(1H,1H,2H,2H-perfluorooctyl) silane vapor (Sigma-Aldrich cas. 78560-45-9). For a ~5 mm-thick device, a PDMS mixture of A and B in 15:1 ratio was poured on the mold, and the bubbles were degassed in a vacuum chamber. After curing the mold at 75°C for 4 h, the PDMS layer was peeled off and the devices were cut. The inlet and outlet holes were punched with 19 gauge needles (McMaster-Carr). All devices were bonded onto a cover glass by plasma bonding. For each experiment, the microfluidic device was filled with S Basal solution to degas. About 100-150 animals were suspended in 1 mL of S Basal and loaded into a syringe. Animals were manually delivered into the device at a flow rate of 2-4 mL h-1. Once the animals were loaded and occupied the individual traps, 1 mL of tetramisole hydrochloride (200 mM) (Sigma-Aldrich cas. 5086-74-8) solution prepared in S Basal was delivered into the device to immobilize the animals. Imaging of the animals in the device was done on a spinning disk confocal microscope (PerkinElmer UltraVIEW VoX). Images were taken were taken with two objectives (10x and 40x), and emission lights were measured with a Hamamatsu

FLASH 4 sCMOS version 2 spinning disk detector. For GFP images, 488 nm excitation laser line and a 527/55 nm bandpass emission filter was used. For RFP images, 561 nm excitation laser line and 615/W70 nm half-power bandwidth emission filter was used. Images of the animals were quantified using ImageJ. A region-of-interest (ROI) was drawn around the entire worm, and the mean intensity of the GFP and mCherry images were calculated across the ROI. Relative fluorescence intensity was calculated as (Mean Intensity of mCherry)/(Mean Intensity of GFP).

Food consumption assay

The experimental method was described in the previous research[74]. In brief, the food consumption assay was performed using 24-well plates seeded with 20μL of freshly-cultured OD600 of 4.0 (CFU ~ 3.2×10^9/mL) *E. coli* OP50-GFP(pFPV25.1). The fluorescence signal of OP50-GFP was quantified by area scanning protocol using BioTek Synergy H4 multimode plate reader. The synchronized L4 animals were placed in the wells in the first 5 columns and the rest column is used as control column. Each well was placed with 10 animals and the plate was incubated at 20°C incubator for 18 hours and the fluorescence signal were quantified again as the ending time point. The relative food consumption amount was calculated using the equations in the previous research[74].

High-throughput growth rates and brood size analysis

The high-throughput growth rates and brood size assays were performed as described previously[163]. In short, approximately 25 bleach-synchronized embryos were aliquoted into

each well of 96-well plates, and fed 5 mg/mL HB101 bacterial lysate on the following day. After 48 hours of growing at 20ºC, a large-particle flow cytometer (COPAS BIOSORT, Union Biometrica, Holliston, MA) was used to sort three L4 larvae into each well of a 96-well plate with 50 µL of K medium with HB101 lysate (10 mg/mL) and Kanamycin (50 µM). Animals were grown for 96 hours at 20ºC and were then treated with sodium azide (50 mM in M9). Animal number (n) and animal length (time of flight, TOF) were measured by the BIOSORT. For each well, animal growth was measured as the median-TOF value of a population, and brood size was measured as the number of progeny per sorted animal. The experiments were replicated in two independent assays, and the linear model with the formula (phenotype ~ assay) was applied to normalize the differences among assays.

CHAPTER 4. STUDY OF GENE-ENVIRONMENT INTERACTION THAT REGULATE *C. ELEGANS* FITNESS

Yuehui Zhao[1], Ed Large[1], Wen Xu[1], Kathryn S Evans[2], Steffen R Hahnel[2], Lijiang Long[1], Erik Andersen[2], and Patrick T. McGrath[1,3,4]

[1]Department of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA

[2]Department of Molecular Biosciences, Northwestern University, Evanston, IL, 60208, USA

[3]Department of Physics, Georgia Institute of Technology, Atlanta, GA 30332, USA

[4]Institute of Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, GA 30332, USA

4.1 Abstract

Genomic variations interact with the environment drive the phenotypic diversity and build the tree of life with vast biological diversity during evolution. Studying the genetic basis of the phenotypic variations and identifying the adaptive mutations can help us understand the mechanism that regulates the biological traits including many human diseases such as cancer and pathogen infection. Characterizing the adaptive mutations and quantify the fitness benefits under different environment is non-trivial. Bringing evolution to the lab and performing the evolutionary experiment provides a route to deal with this challenge. In this chapter, I utilized pairwise competition experiment to identify the adaptive alleles and quantify their fitness effects under different environments including three different feeding conditions, two different temperatures, and the growth condition with anthelmintic drugs. This work demonstrates that the evolutionary experiment is a powerful tool to study how gene-environment interaction regulate fitness.

4.2 Introduction

There is a general interest to study how the adaptive mutations interact with the environment and lead to the fitness benefits to the organisms. For instance, many genetic variations in the human population were proposed to facilitate human to adapt to new diets [164-166]. One example is the allele in lactase. The domestication of livestock provided human access to milk, but lactase, the enzyme necessary to metabolize lactose sugar, is shutoff in adult mammals. The derived alleles that result in the persistence of this enzyme in adults, and consequently metabolism of lactose result in fitness benefits to human population[167]. Nowadays, with the development of modern agricultural techniques, many of us are exposed to new types of diets with increased amounts of sugar and fats that can lead to obesity and can further cause heart problem, Type 2 diabetes, and some types of cancer. Thus, identification of the adaptive mutations and study how they interact with the environment can be important to help us to adjust our living habit to prevent disease, and give us insights into the potential drug targets for developing a new treatment strategy. Many interesting biological questions could also be more investigated: the organisms adapt to the habitat with different climate conditions? How the pathogenic organisms derive the drug resistance phenotype?

In this chapter, I will describe the researches that apply the pair-wise competition experiment to characterize the adaptive mutations in two laboratory domesticated strains N2 and LSJ2 under different feeding conditions: 1) normal growth condition; 2) starving growth condition; 3) axenic HS-YE-HLE liquid media. N2 and LSJ2 share the same

ancestor but were domesticated in two different growth media that serve different food source. During the decades of adaptation, there are ~300 genetic variants between the two lineages. This provides us a good source to study which derived mutations facilitate them to adapt to the new diets. The interesting biological questions could be studied like whether the adapted mutations increase fitness in one diet can also show fitness benefits in the other diet source. In addition, I will also describe the two collaboration projects that 1) study how the different temperatures affect the fitness of two wild isolate *C. elegans* strains[26]; 2) quantify the fitness effect of the alleles that resistant to anthelmintic drugs.

4.3 Results

4.3.1 Fitness QTL mapping identify a single significant QTL in normal growth condition

In Chapter 3, I have performed the pairwise competition experiment and measure the relative fitness of each CX12311(N2*) x LSJ2 RIL. This dataset could be used to identify if there are other derived N2 alleles or LSJ2 alleles are adaptive in addition to *npr-1* and *glb*-5. The distribution of the relative fitness shows the bimodal distribution that suggests this is a mendelian trait regulated by one genetic locus. By performing the QTL analysis using 192 markers of N2 and LSJ2 genetic variants, a significant QTL was found at the right arm of Chromosome II around 14 Mbps which close to the chromatin remodeling factor *nurf-1* (**Figure 4.3.1a**). *nurf-1* is orthologous to the human gene that encodes the BPTF subunit of NURF complex. In LSJ2, there is a 60bp deletion at *nurf-1* C-terminus which causes the truncation of the long isoforms of *nurf-1* gene. To identify if this 60bp deletion is a casual allele that affects fitness, the *nurf-1* NIL strain PTM66 (NIL_{nurf-1}) carries the LSJ2 region including 60bp deletion allele in CX12311 background and the CRISPR/Cas9 engineered allelic replacement strain PTM88 (ARL_{nurf-1}) that carries the LSJ2 60bp deletion allele were used for pairwise competition experiment. The result indicates that the 60bp deletion at *nurf-1* gene significantly decreases fitness (**Figure 4.3.1b**). In addition, NIL_{nurf-1} and ARL_{nurf-1} show the significant difference which suggests there are the other alleles close to *nurf-1* 60bp deletion affect fitness effect.

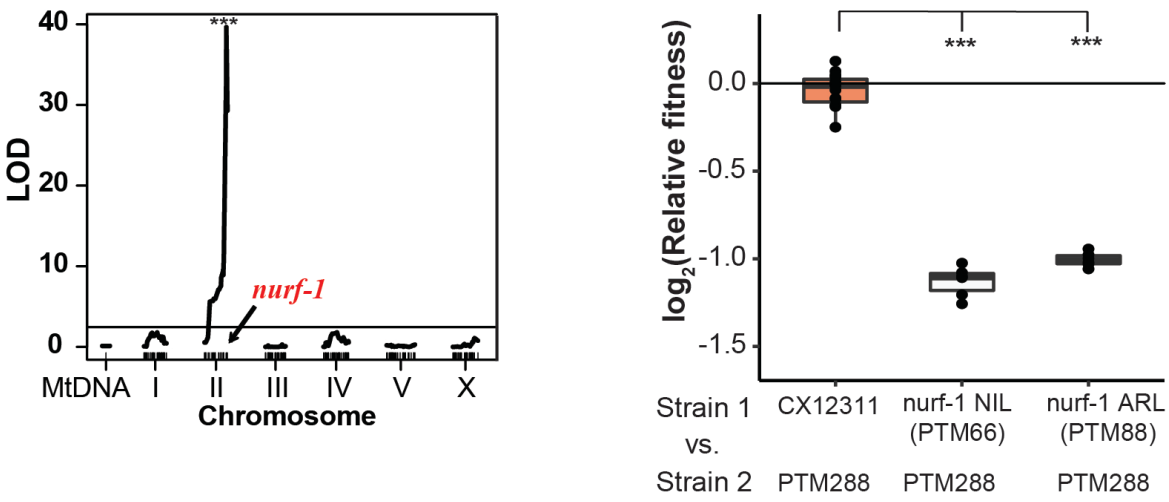

Figure 4.3.1. *nurf-1* is a major factor to regulate fitness in normal growth condition. **a.** Fitness QTL analysis of CX12311(N2*)/LSJ2 RILs in normal growth condition. The black threshold line shows the significance level $\alpha = 0.05$ in 1,000 permutation experiment for one-dimensional scan. **b.** Pairwise competition experiment shows LSJ2 derived 60bp deletion decrease fitness in normal growth condition. ***p<0.001 by ANOVA with Tukey's Honest Significant Difference test.

4.3.2 LSJ2 derived nurf-1 allele show opposite fitness effects in two growth media

The allele of 60bp deletion at *nurf-1* gene decrease the fitness in standard growth condition in the agar plate. This allele is fixed in LSJ2, which domesticated in HS-YE_HLE media around 50 years. It is interesting to study its fitness effect in liquid media to investigate if this allele might be fixed by either genetic drift or natural selection. To answer this question, the pairwise competition experiment was performed in the agar plate and liquid media. The result indicates that although LSJ2 derived allele of 60bp deletion is deleterious in agar plate growth condition, it shows fitness benefit in liquid media (**Figure 4.3.2a**). The LSJ2

derived *nurf-1* allele also shows smaller size and delayed egg laying timing compare to N2 strain. These life-history traits changes inspire us to study if the *nurf-1* allele affects longevity. The lifespan assay was performed using the CX12311, LSJ2, MT13694, and ARL_{nurf-1}. MT13694 carries the 1078bp deletion at *nurf-1* 3' end. The lifespan assay indicates that 60bp deletion significantly extends the animal's longevity (**Figure 4.3.2b**).

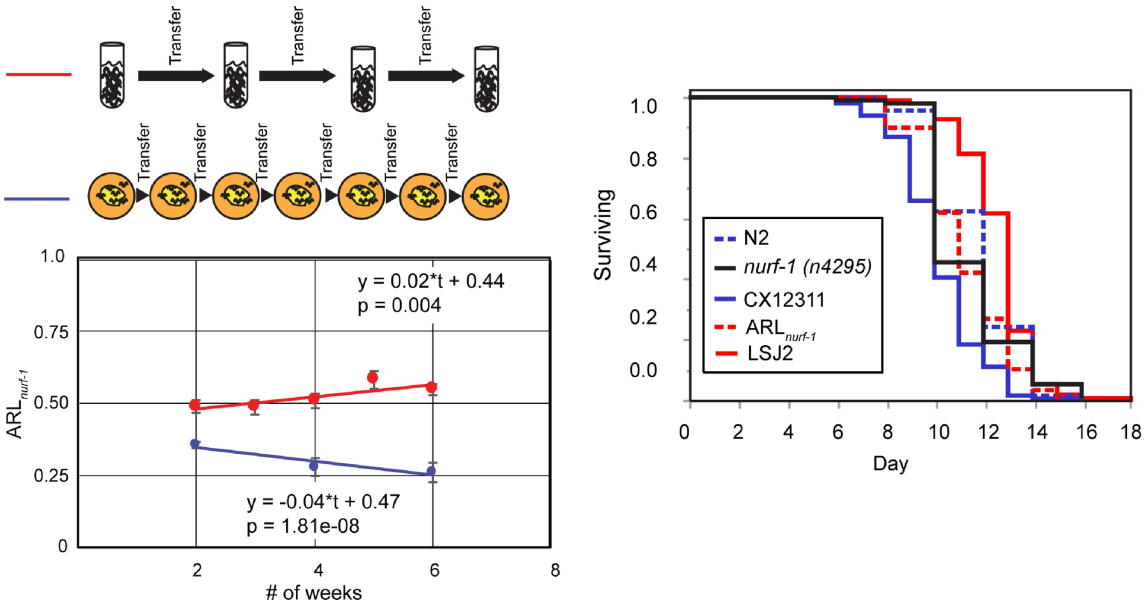

Figure 4.3.2. LSJ2 derived allele of *nurf-1* show opposite fitness effect in two growth media and extend longevity in agar plate condition. a. Pairwise competition experiment of ARL_{nurf-1} vs. CX12311 in an agar plate and liquid media. The diagram of the two competition conditions is shown above the graph. The ARL_{nurf-1} strain (PTM88) competed against CX12311. The fraction of ARL_{nurf-1} animals in the population is shown on the y-axis. The equation for a linear regression fit is also shown on the graph. p-value indicates the significance of the temporal term by ANOVA. Error bars represent standard error. **b.** Lifespan analysis of the N2, LSJ2, CX12311, $ARL_{nurf-1,}$ and *nurf-1(4295)* animals. At least three independent replicates containing ~60 worms were used for this analysis. The LSJ2

and $ARL_{nurf\text{-}1}$ strain were both significantly different ($p < 0.05$) than CX12311. The *nurf-1(4295)* strain was not significantly different from the N2 strain.

4.3.3 G × G and G × E were found to regulate fitness in normal and starve growth condition

The N2 lineage was cultivated in agar plate for around 20 years. Although the animals were served with sufficient food source at the beginning, their growth in the agar plate could encounter starve state when the food was consumed by rapidly increased population and delayed transfer. When *C. elegans* undergo the lacking of food supply or stressful environments such as high population density or increasing temperature, the animals will adopt an alternative developmental strategy that allows L2 stage worms to enter into dauer stage. After change the growth environment with a food source and the stress factors were ruled out, the dauer will go back to the normal developmental cycle. To study if there are any derived mutations contribute to the fitness benefits under starving states, the modified competition experiments were carried out to allow maintenance of animals for two more weeks after all the food were consumed. The relative fitness of each RIL without serious fungal contamination was quantified (**Figure 4.3.3**).

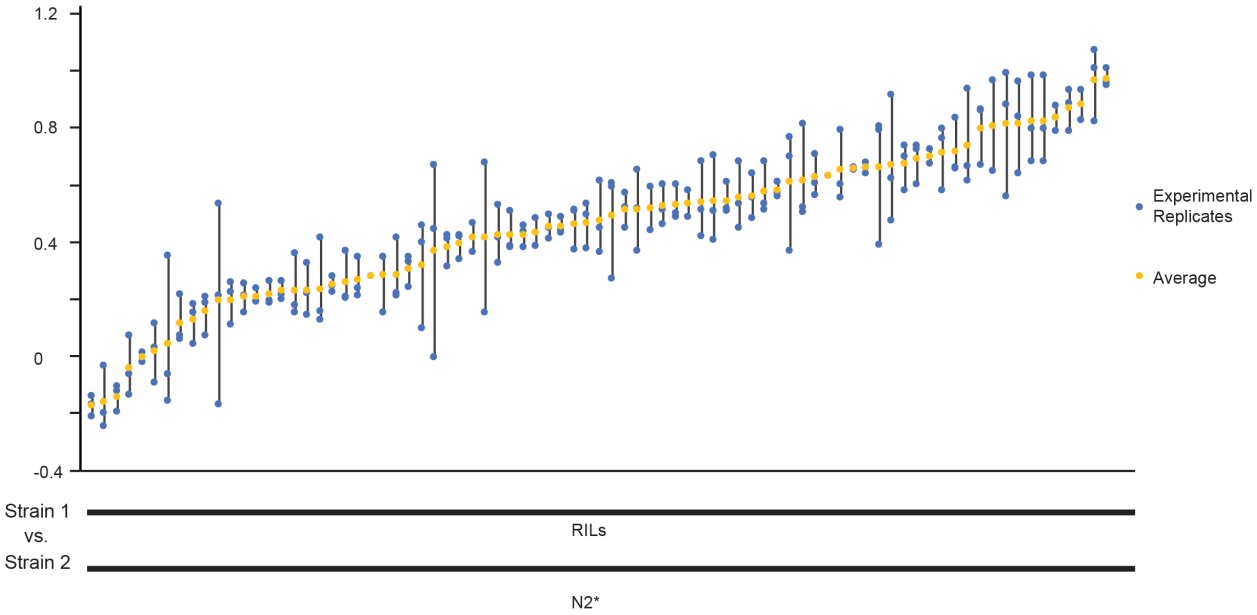

Figure 4.3.3. Relative fitness distribution of the N2*(CX12311)/LSJ2 RILs in starve growth condition. Pairwise competition experiment of each N2*(CX12311)/LSJ2 RILs vs. PTM288 in starve growth condition in the agar plates. The blue dots represent the relative fitness data of experimental replicates and the yellow dots represent the average value of relative fitness of each RIL.

The single QTL at *nurf-1* locus was found by QTL mapping (**Figure 4.3.4a**). By subtracting the relative fitness data in normal growth and the data in starve growth condition, the QTL mapping was performed using the data of relative fitness difference between two feeding conditions. Two more QTL were found at ChrV ~4,5Mbps including the whole *Y61A9LA.3* gene deletion and ChrX ~14Mbps including the G-protein coupled receptors *srg-36 srg-37* deletion (**Figure 4.3.4b**). Using *nurf-1* marker as a covariate to scan the loci show interaction with *nurf-1*, the ChrV and ChrX show statistical significant that interact with *nurf-1* (**Figure 4.3.4c**).

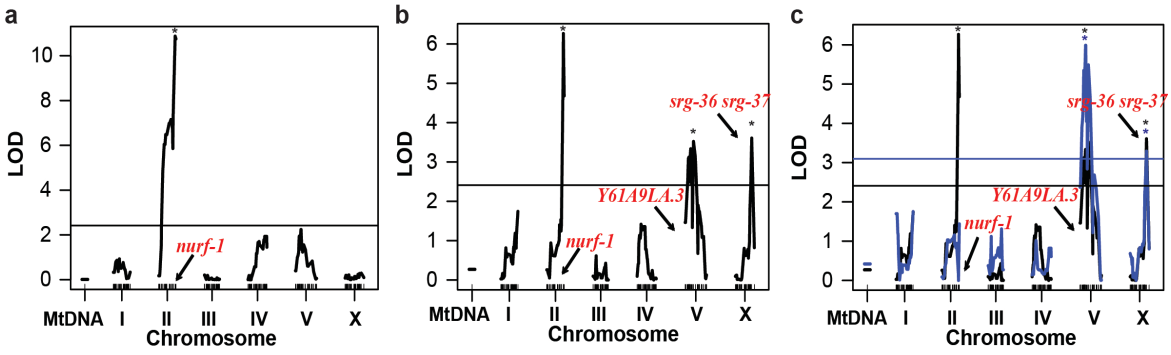

Figure 4.3.4. Fitness QTL analysis of starve growth group and the difference between normal growth and starve growth conditions. a. Fitness QTL analysis of the starve growth group. **b.** Fitness QTL analysis of the difference between normal growth and starve growth conditions. **c.** Fitness QTL analysis that mapping the QTL show interaction with *nurf-1*. The black threshold line shows the significance level $\alpha = 0.05$ in 1,000 permutation experiment for one-dimensional scan. The blue threshold line shows the significance level $\alpha = 0.05$ in 1,000 permutation experiment for one-dimensional scan with a *nurf-1* locus as a covariate.

The LSJ2 alleles at ChrV QTL do not show fitness effect under N2 *nurf-1* background, while decrease fitness under LSJ2 *nurf-1* background in normal growth condition. However, the LSJ2 alleles at ChrV QTL do increase fitness effect under N2 *nurf-1* background and do not affect fitness under LSJ2 *nurf-1* background (**Figure 4.3.5a**). This result found the G x G interactions between *nurf-1* and the alleles at ChrV QTL as well as a strong G x E interactions between the two feeding conditions. In addition, another QTL at ChrX also shows G x E interactions in a genetic additive effect manner with *nurf-1*. This QTL including *srg-36* and *srg-37* deletion, which are necessary for pheromone sensing and dauer formation. The result shows that the LSJ2 alleles at ChrX QTL increase fitness in

normal growth condition while decrease fitness under both N2 and LSJ2 *nurf-1* background in starve growth condition (**Figure 4.3.5b**).

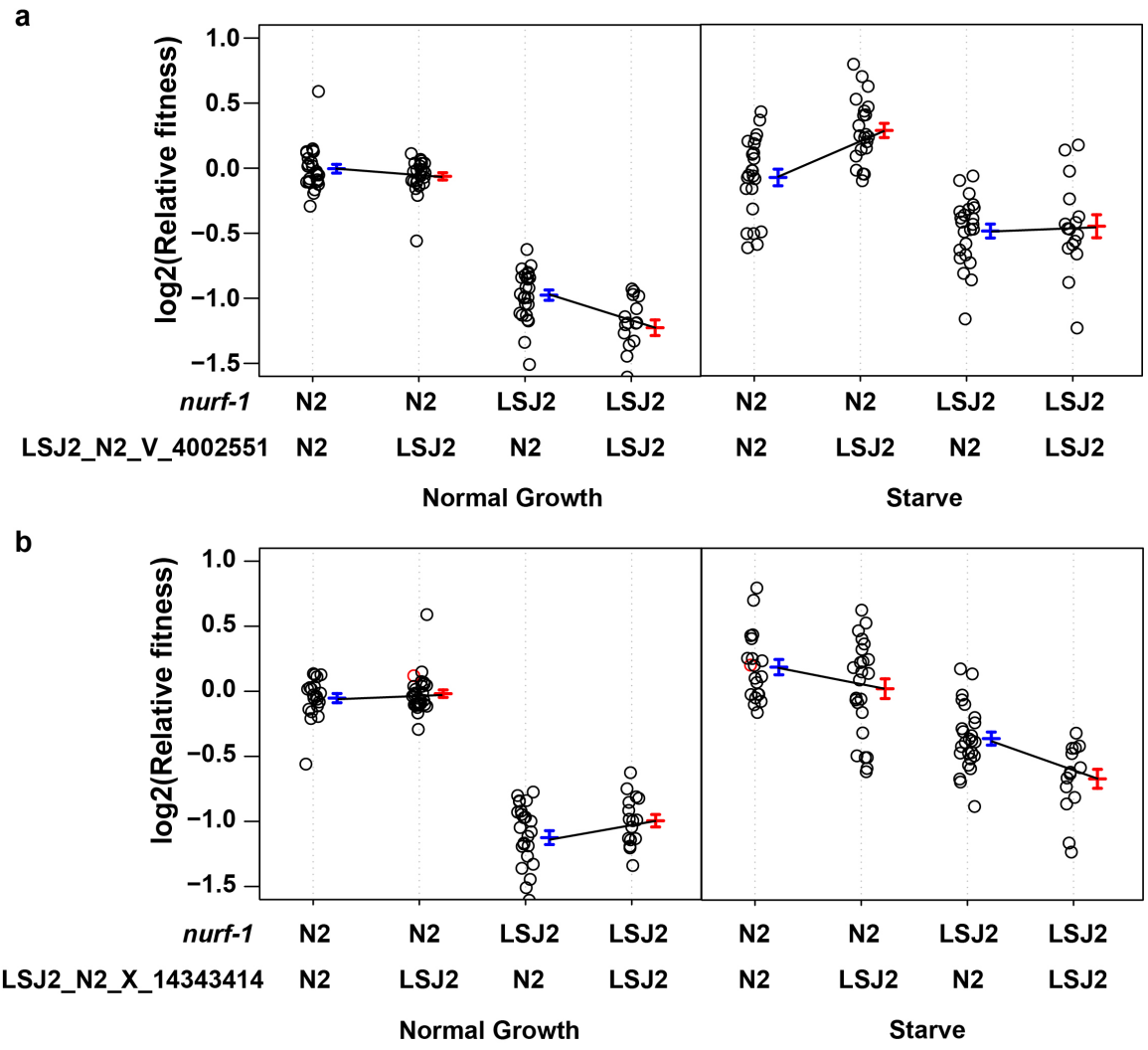

Figure 4.3.5. GxG and GxE effect of the ChrV and ChrX QTL in the normal growth condition and the starve growth condition. a. GxG and GxE effect of ChrV QTL marker (ChrV 4,002,551) and *nurf-1* marker. **b.** GxG and GxE effect of ChrX QTL marker (ChrX 14,343,414) and *nurf-1* marker.

4.3.4 Fitness QTL analysis in liquid media

LSJ2 lineage was continually domesticated in HS-YE_HLE media for ~ 50 years and was frozen in 2009. Before LSJ2 lineage was frozen, the other lineage LSJ1 was frozen in 1995. To systematically identify the adaptive mutations that have fitness benefits in liquid media. I carried out the pairwise competition experiment that used CX12311 x LSJ2 RILs, as well as LSJ1 and LSJ2, compete with PTM288 strain (N2*_bc) that carries *dpy-10(kah83)* silent mutation in CX12311 background. Both LSJ1 and LSJ2 show higher fitness than PTM288 in liquid media (**Figure 4.3.6a**). The distribution of the relative fitness shows a continuous profile that suggests the fitness benefits in LSJ2 strain has more complex genetic basis (**4.3.6b**). By applying QTL analysis, a significant QTL was found in ChrX around 5.7Mbps with relative lower LOD score ≈ 3 (**4.3.6c**). There are two intergenic variants next to this QTL. One of these two variants may have a modifier effect on *F46G11.2* gene. The results might imply there are multiple mutations show fitness benefits in LSJ2 lineage that across all the chromosomes.

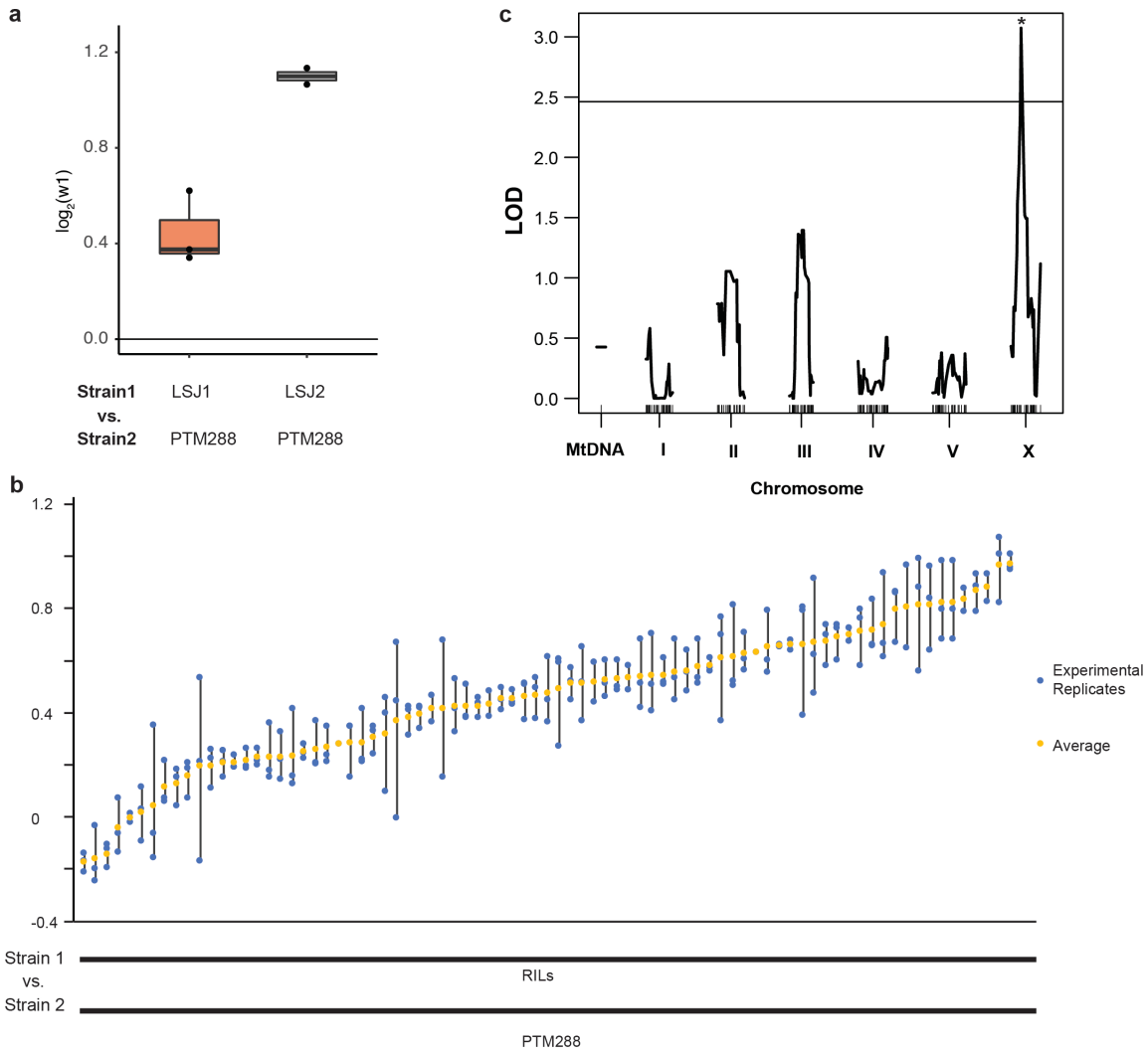

Figure 4.3.6. High-throughput fitness assay of CX12311/LSJ2 RILs in liquid media. **a.** Pairwise competition experiment of LSJ1/LSJ2 vs. PTM288. **b.** Pairwise competition experiment of each CX12311/LSJ2 RILs vs. PTM288 in liquid media. The blue dots represent the relative fitness data of experimental replicates and the yellow dots represent the average of the value of relative fitness of each RIL. **c.** Fitness QTL analysis of liquid competition group. The threshold line shows the significance level $\alpha = 0.05$ in 1,000 permutation experiment.

4.3.5 QTL associated with temperature parameter shows fitness effect difference at two temperature conditions

To study how the organism adapts to a new climate condition can give us valuable information to encounter global climate changes. By collaborating with Dr. Erik Andersen's group at Northwestern University, we used two wild isolates JU847 and CX11314 to study their fitness difference in two different temperature conditions. These two strains were isolated from locations with different climate parameters. JU847 was isolated from Northern France in 2005 with a three-year average temperature of 11.3 ^{o}C. CX11314 was isolated from Southern California with a three-year average temperature of 20.9 ^{o}C. These two strains carry the different alleles at a ChrV QTL show strong association with temperature parameter that identified by the GWA study[26]. To determine if they show relative fitness difference in low and high-temperature condition, the pairwise competition experiment between these two strains was performed at 15 ^{o}C and 25 ^{o}C. CX11314 was found to have higher fitness than JU847 at both temperatures tested (**Figure 4.3.7**). At high temperature, CX11314 had a clear selective advantage compared to JU847 (for fitness = 1, relative CX11314 fitness s = 2.29), resulting in JU847 alleles comprising fewer than 1% of the alleles measured after six culture transfers. However, JU847 performed better at the lower temperature than at a higher temperature, comprising almost 8% of the total nematode population after six culture transfers (relative CX11314 fitness s = 1.57). These data suggest that JU847, although not more fit than CX11314 at either temperature, is more fit at 15° than at 25°.

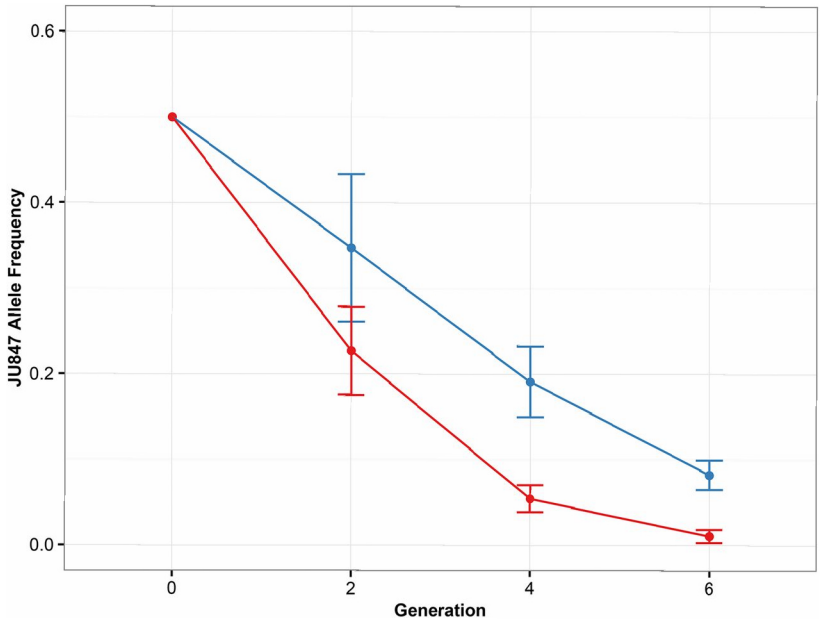

Figure 4.3.7. Temperature competition assay of JU847 and CX11314. JU847 (isolated at low temperature) was competed against CX11314 (isolated at high temperature) at both 15° C (indicated in blue) and 25°C (indicated in red). The mean frequency of the JU847 allele in the population is plotted on the y-axis. Error bars represent 1 SD from the mean. Data were collected from nine experimental and five technical replicates.

4.3.6 Quantification fitness effect of an anthelmintic drug resistance allele in C. elegans

To study how the pathogens evolve the drug resistance is important for effective prevention and treatment of the infective diseases. The nematode infections threaten to a large human population. The control of nematode infections is mainly relied on the limited anthelmintic drugs including the benzimidazoles(BZ). Dr. Erik Andersen's group at Northwestern University has identified the genetic basis of *C. elegans* resistance to BZ and albendazole (ABZ) by genome-wide association mappings[168]. They found *ben-1* gene, encodes the β-

tubulin is a major determinant of ABZ resistance. One of the variants is *ben-1* F200Y. By collaborating with Dr. Erik Andersen's group, we are quantifying the fitness effect of *ben-1* deletion and F200Y variant under the treatment with ABZ. In the competition experiments, we individually competed the F200Y, Del, and parental N2 strains against an N2 strain that contains a barcode sequence (PTM229). We quantified the relative allele frequencies of the barcoded strain for the first, third, fifth, and seventh generations of the competition assay (**Fig 4.3.8**). Throughout the competition assay, and for all strains tested, the relative frequencies of the barcoded strain did not significantly deviate from the initial frequency when grown on DMSO plates. These results suggest that in standard laboratory conditions, the BEN-1 F200Y and ben-1 deletion alleles do not have fitness consequences. We observed the same trend when we competed for the N2 and the barcoded N2 strains on ABZ plates. However, the allele frequencies of the barcoded strain dropped to ~20% when competed against strains that contain either of the two ben-1 alleles on ABZ plates (relative fitness w = ~ 1.3, both). These two independent assays show that the two ben-1 alleles confer BZ resistance with no negative fitness consequence under standard laboratory growth conditions.

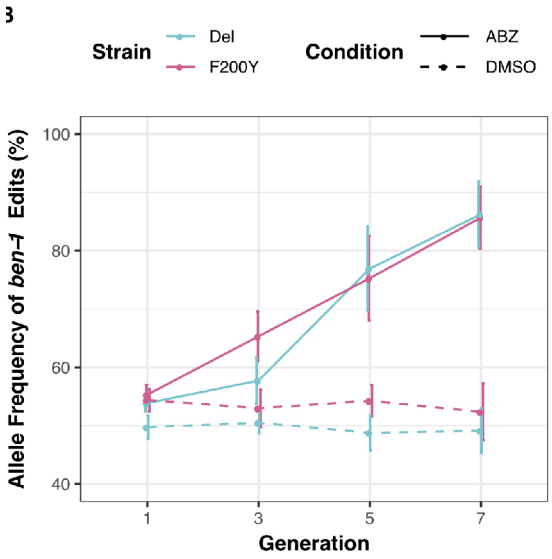

Figure 4.3.8. The competition experiment of the strains carry *ben-1* deletion and *F200Y* under ABZ treatment. The results from a multi-generation competition experiment between a barcoded N2 strain (PTM229) and the ben-1 allele-replacement strains are shown. The generation number is shown on the x-axis, and the allele frequency of the ben-1 allele-replacement strains is shown on the y-axis.

4.4 Discussion

From the fitness QTL analysis in normal growth condition, I have shown that *nurf-1* plays a major role to regulate fitness effect in CX12311/LSJ2 RILs. The LSJ2 derived 60bp deletion is close to the Bromo domain and PHD domain that binds to histone modifications H3K4 tri-methylation and H4K16 acetylation[169]. This 60bp deletion may affect the function of those two domains and further affect chromatin remodeling and gene expression. It is intriguing to see that NURF chromatin remodeling factor affects multiple including egg-laying rates delay and extended longevity. Human BPTF was reported involved in cancer process[170]. This finding may inspire other researchers to study the if NURF remodeling factor affects germline cell fate during differentiation and cell cycle control process.

By comparing the RILs fitness in two different feeding conditions, I have shown two more QTL at ChrV and ChrX show GxG and GxE effects to regulate relative fitness. ChrV QTL including the whole *Y61A9LA.3* gene deletion. This gene encodes multiple isoforms but with very limited knowledge. The deletion of this gene shows fitness increase in starve growth condition and suppresses the deleterious fitness effect of *nurf-1* 60bp deletion. It will be intriguing to study its function in dauer formation and recovery as well as how it interacts with *nurf-1* to regulate the fitness effect. The ChrX QTL including the deletion of *srg-36* and *srg-37*. Loss of functions of *srg-36* and *srg-37* increase fitness in normal growth condition but decrease fitness in starve growth condition. The potential reason explains this is these two genes play a role to sense dauer pheromone and contribute to dauer formation. The dauer stage is the specific worm stage facilitates the animals to survive in the stressful

condition including limited food sources. In the normal growth condition with plenty of food source, expression of these two genes can lead to sensitivity to sense the intense population around and result in slower development and decreased number of egg laying. Thus, loss of function allele could lead to fitness increase in this condition. However, these two genes play important roles in dauer formation. Once the food source is running out, the animals will enter into the dauer stage for survival. Deletion of these two genes may affect dauer formation which can lead to a deleterious effect on fitness.

In the high-throughput fitness assays in liquid media, there is one single significant QTL was identified with relative lower LOD compare to *nurf-1* QTL, ChrV QTL, and ChrX QTL. This result suggests that multiple LSJ2 derived alleles contribute to fitness benefits in liquid media. In addition, we need to consider the factors in the experiments. Performing the competition experiment in liquid media is non-trivial. The nutrient components of the beef livers gained from the pet food store and the HY-soy peptones may not be identical for each batch. Since lack of the record how the previous researchers grow the strains in liquid media, my experimental condition is not identical to the original LSJ2 domestication environment. These variables should be also considered. In this part of the research, I will suggest the follower researchers redo a small number of RILs (6-10) as well as LSJ1 and LSJ2 to study the potential variability from different batches of beef liver and HY-soy peptones.

In the last two results of this chapter, I demonstrate the competition experiment I developed to show its power to quantify the fitness effect of genetic variants with interest in different

environmental conditions. Applying this approach can also provide a potential route to characterize the fitness effect of the drug resistance mutations in other pathogenic organisms including bacteria and fungal. It may also be applied to study the chemoresistance in human cancer clones which can give doctors more clues to improve the cancer treatment strategy.

4.5 Materials and Methods

Nematode Growth Conditions

The strains used for the experiments in agar plates growth condition

The animals were cultivated on standard nematode growth medium (NGM) plates containing 2% agar seeded with 200 µl of an overnight culture of the *E. coli* strain OP50 in an incubator set at 20°C. Strains were grown for at least three generations without starvation before any experiments were conducted.

The strains used for the experiments in liquid media growth condition

The animals were cultivated in 10mL HS-YE-HLE stock media at 20°C with 100 rpm shaking on an orbital shaker or 10mL HS-YE-HLE stock media in an incubator at 20°C. Strains were grown for at least 2 weeks before the competition experiments were conducted.

Strains

Laboratory strains and wild isolates

N2, LSJ1, LSJ2, LSJ3, CX11314, JU847

NIL strains

CX12311 (N2*) - *kyIR1(V, CB4856>N2), qgIR1(X, CB4856>N2)*; PTM66 - *kyIR87(II, LSJ2>N2); kyIR1(V, CB4856>N2); qgIR1(X, CB4856>N2)*

RIL strains

CX12311 – LSJ2 RILs: CX12312-19, CX12321-27, CX12346-52, CX12354-60, CX12362-66, CX12368-75, CX12381-88, CX12414-37, CX12495-99, CX12501-08, CX12510, CX12361

Mutants

PTM88 - *kyIR1 (V, CB4856>N2), qgIR1 (X, CB4856>N2), nurf-1(kah3);* PTM229 - *dpy-10(kah81);* ECA884 – *ben-1(ean66);* ECA917 – *ben-1*(*ean918*); MT13649 *nurf-1 (n4295).*

HS-YE-HLE stock media preparation

The axenic liquid HS-YE-HLE stock media was prepared by mixing one volume of Heated Liver Extract (HLE) to nine volumes of HySoy-Yeast Extract (HS-YE). The preparation method is modified from the following paper[171]. To make the HLE component, calf liver (purchased from Corrina's Corner) was cut to 1-inch squares and left in 4°C cold room overnight for 24 hours. An equal amount of distilled water was added to the liver, which was then further broken down using a blender. Large particles were filtered out of the homogenate with Miracloth. The purified homogenate was then heated in a 60°C water bath until its temperature reached 52°C and heated for 6 minutes. At that point the homogenate was centrifuge and the supernatant further filtered with a 0.2-micron filter (Nalgene). To prepare the HS-YE component, 40 g HySoy peptone (Sigma P6463) and 10 g yeast extract (Alfa Aesar H26769) was added to 1 L of water and autoclaved. The final HS-YE-HLE stock media also contained three antibiotics (Penicillin G 100 U/mL;

Streptomycin 100 ug/mL and Amphotericin B 0.25 ug/mLand was filtered with a 0.2-micron filter (Nalgene).

Competition experiment in different environmental conditions

Competition experiment to study the fitness effect of a derived allele of *nurf-1*

Competition assays between CX12311 and ARLnurf-1 (PTM88) were performed on NGM plates and liquid HS-YE-HLE stock media. An SNV located on chromosome I at position 11583395 (WS220) (CX12311: T, *ARLnurf-1*: C) was used to genotype and quantify the number of each strain in the competition experiments. In the competition assay on NGM plates seeded with OP50 bacteria, nine experimental replicates were performed. Each replicate included five independent populations. At the beginning of each assay, the animals were synchronized with alkaline-bleach and raised at 20°C until they developed to the L4 stage. Seven L4 animals of each strain were placed on forty-five 6 cm NGM plates and kept at 20°C for one week. At this point, populations from each plate were transferred to new NGM plates by cutting a 0.5 cm x 0.5 cm square of agar (containing ~100 worms) from the starved plate. Populations were continuously cultured in this way for five weeks. During the six weeks of culturing, genomic DNA was isolated from the 2nd, 4th and 6th week time points. The populations from five plates in each group were combined into a single Eppendorf tube and genomic DNA was isolated by using a Qiagen Gentra Puregene Kit (cat. nos. 158667) following the supplementary protocol for nematodes and purified using Zymo Quick-DNA universal kit (cat. nos. D4068). In the competition assay in liquid HS-YE-HLE stock media, six experimental replicates were performed. Each replicate was

cultured in a 10 mL HS-YE-HLE stock media in a cell culture flask. At the beginning of this assay, the worms were synchronized by alkaline-bleach. When the animals were hatched, 100 L1 animals from each line were transferred into 10 mL of HS-YE-HLE media and cultured at 20°C in a vertical shaker. After two weeks, 1 mL of depleted culture (containing ~2000 animals) were transferred to a new cell culture flask container. Then, the populations were continuously cultured for another four weeks. The populations were transferred every two weeks and the populations' genomic DNA was isolated at the 2nd, 3rd, 4th, 5th and 6th-week time point by the same method described above. The proportion of ARLnurf-1 population size in a competition assay was measured using Taqman analysis in Biorad QX200 digital PCR system to quantify the chromosome I 11583395 SNP frequency. TaqMan probes were designed using standard software from Applied Biosystems. Genomic DNA from each time point was digested with SacI enzyme and purified with Zymo DNA Clean & Concentrator Kit (cas. nos. D4004). The concentration of fragmented genomic DNA was adjusted to 2 ng/uL by Qubit assay (cas. nos. Q32851). Digital PCR was performed followed the standard method provided by Biorad with the absolute quantification method. The proportion of ARL nurf-1 allele was calculated and the linear model and statistical analysis were carried out in the R programming language.

Competition experiment under different temperature

We chose two strains, CX11314 and JU847, that had different alleles for the peak QTL marker (chrV: 14,822,276; JU847: T, CX11314: A) in our three-year temperature GWA mapping. JU847 has the reference allele for the peak marker and was isolated at a low temperature, whereas CX11314 has the alternative allele for the peak marker and was isolated at a higher temperature. We designed a TaqMan probe (5′-

[A]CCGTTTTTTTT[T/A]AATTTT-3′) to measure each of these two alleles from mixed samples of nematodes using the standard software from Applied Biosystems (https://www.thermofisher.com/order/custom-genomic-products/tools/genotyping/) and a corresponding primer set to amplify the region of interest (below).

F: 5′-AAACCCAAGATTTTTATGGTTACTTTAAGATTTGT-3′;

R: 5′-ATCTATAGTTAACTTGGATATATTGTTTGTTTTCGGT-3′

These two strains were chunked to fresh 10 cm NGMA plates seeded with OP50; 48 hr later, seven L4s from each strain were added to each of 45 6-cm NGMA plates seeded with OP50 for both 15°C and 25°C competition experiments. The 45 plates at each temperature represent nine experimental replicates, each composed of five independent populations. Plates were placed at either 15 or 25°, and grown to starvation. After one week for 25° competition experiments, or 10 d for 15° competition experiments, nematodes were transferred to fresh NGMA plates by cutting a 0.5 × 0.5 cm square of agar (containing ~100 worms) and replaced at the appropriate temperature. After culture transfers, two, four, and six, starved animals were washed off the plates with M9, and DNA was collected using the Qiagen DNeasy Kit. Genomic DNA from each time point was digested with the EcoRI enzyme and purified using the Zymo DNA Clean and Concentrator Kit. The concentration of fragmented genomic DNA was adjusted to 2 ng/µl by Qubit assay. The number of JU847 and CX11314 alleles in each replicate population was measured using Taqman analysis in a Bio-Rad QZ200 digital droplet PCR system. Digital PCR was performed following the standard protocol provided by Bio-Rad with the absolute quantification method. The proportion of the JU847 allele and the relative selection coefficients were calculated.

Competition experiment under different feeding conditions

Standard growth condition: The method and the data of the competition experiment using CX12311(N2*) - LSJ2 RILs under standard growth condition is the same as the high-throughput competition experiment (Figure 3.2.2). The starving growth condition: the setup of strains on the 1st competition plate is the same as normal growth competition. After 5 days, all the foods were consumed. 2mL M9 buffer was added to the agar plate and 1mL of M9 with worms were transferred to 1.5mL centrifuge tubes for the following genomic DNA isolation and droplet digital PCR assay. The rest population on the plate was maintained for 16 more days in total three weeks in the plate running out of food, 1cm x 1cm agar chunk including animals in dauer state was cut by spatula and transferred to the new plate. Before the next transfer, half of the animals were washed off on Day 4 when foods were running out for quantification of the allele frequency. The rest of the animals were maintained for another 17 days in total 3 weeks for the next transfer to the next plate. In total, there are two transfers and the allele frequency in the population was quantified from the three time-series plates. The relative fitness of each RIL was quantified using the same method described in chapter 3. The liquid growth condition: 100 L1-L3 stage animals were seeded into 10mL HY-YE-HLE liquid media and cultured at 20°C in the incubator. The animals were cultured in the media for three weeks. 1mL culture was transferred into 1.5mL centrifuge tubes for genomic DNA isolation and the allele frequency quantification. 200uL culture was transferred into a new 10mL HY-YE-HLE liquid media. In total, there are three two transfers and the allele frequency in the population was quantified from the tree time-series liquid media culture. The relative fitness of each RIL was quantified using the same method described in chapter 3.

Competition experiment under treatment with albendazole (ABZ)

Pairwise multi-generation competition assays were performed between a ben-1 wild-type strain PTM229 and the ben-1 edited strains, containing either the F200Y allele replacement or a ben-1 deletion. All strains were generated in the N2 background. For the assay, strains were bleach-synchronized and embryos were transferred to 10 cm NGM plates. 48 hours later, seven L4 larvae per strain were transferred to 50 fresh 6 cm NGM plates containing either 1.25 μM ABZ for drug selection or DMSO as a control. The ABZ concentration was determined by dose-response assays beforehand as the lowest concentration that caused a developmental delay in PTM229 and N2 when added to NGM. For each strain combination, 50 plates were grown, representing 10 technical replicates of five independent populations. Plates were grown for one week until starvation. Animals were transferred to fresh plates by cutting out a 0.5 cm3 agar chunk. After every culture transfer, starved animals were washed off the plates with M9, and DNA was collected using the Qiagen DNeasy Kit (Catalog #69506). Allele frequencies of PTM229 compared to wild-type N2 or the ben-1 edit strains in each replicate populations were measured using Taqman analysis in a Bio-Rad QX200 digital droplet PCR system. Digital PCR was performed following the standard protocol provided by Bio-Rad with the absolute quantification method. To calculate the relative fitness w of the competitive strains, we used linear regression to fit the relative allele frequencies at the first, third, fifth, and seventh weeks into a one-locus genic selection model[153].

QTL mapping

R/qtl was applied to perform a one-dimensional scan using the marker regression on the 192 CX12311(N2*) x LSJ2 markers[15,71]. The significance threshold (p = 0.05) was determined by 1,000 permutations tests. To identify the loci show interactions with *nurf-1*, the *nurf-1* marker was used as an additive and interactive covariate for one-dimensional scans. The significance threshold (p = 0.05) was determined by 1,000 permutations tests.

Life span assay

Lifespan assays were performed with the standard method at 25°C using NGM plates containing 25 μM FUdR seeded with OP50. The animals were synchronized using alkaline-bleach to isolate embryos and raised on NGM plates at 20°C until they reached the young adult stage when they were transferred to FUdR plates at 25°C. The animal survival number was scored every two days. Animals were scored as dead when they no longer responded to gentle touch with a platinum wire. The date when the animals were placed on FUdR plate was defined at t = 0. The survival statistical analysis was performed in JMP12 software by using the log-rank method in Kaplan-Meier survival.

CHAPTER 5. CONCLUSIONS AND FUTURE STUDIES

My doctoral thesis applied the experimental evolution approach to study how the two laboratory domesticated *C. elegans* strains adapt to the new growth environment and food sources. The two *C. elegans* strains N2 and LSJ2 serve as an unintentional long-term experimental evolution model. Identifying the genetic and phenotypic basis increase the fitness in a different environment can give us many important insights about how metazoan organisms evolve.

The work in Chapter 2 found although the derived alleles of *npr-1* and *glb-5* show fitness effect, their derived solitary behavior does not necessary for fitness advantages. In addition, I also found the oxygen sensing neural circuit URX-RMG play roles in regulating animal's feeding behavior and reproductive timing. This work shows the fitness increase does not due to the derived solitary behavior but by other biological traits. The behavior change is a pleiotropy effect. Applying advanced bioinformatics analysis can interpret which genetic changes are under selection, but it is difficult to study which biological traits regulated by the genetic changes are responsible for fitness benefits. This work demonstrates that applying experimental evolution shows great power to resolve this challenge. There are many interesting questions could be further explored: 1) The previous research identified the oxygen sensing neuron URX can regulate fat metabolism and this work shows its function to regulate food consumption[80,153]. It will be intriguing to characterize the molecular cellular mechanism and dissect the neural circuit that how they regulate the feeding behavior. 2) *daf-22* is essential to fitness benefits of the derived alleles of *npr-1*

and *glb-5*. Pheromones have been shown to influence fat metabolism via ADL sensory neurons[145]. ADL neurons are regulated by pheromones in an *npr-1*-dependent manner[78,128,132,135,136]. It is interesting to do further study if ADL is playing role in the fitness benefits of *npr-1*. 3) Derived *glb-5* is a loss of function mutant. Ablation of URX neuron indicates it shows a deleterious effect on fitness. In addition to expression in URX, AQR, and PQR oxygen sensing neurons, it is also expressed in a couple of other neurons[19]. It could be interesting to investigate the other molecular roles of *glb-5* in another neural circuit.

In Chapter 3, I applied a high-throughput fitness experiment and identified the beneficial complex genomic rearrangements at calcipressin *rcan-1* gene. The *rcan-1* complex genomic rearrangements lead to active exploration behavior and fitness benefits. Although the complex genomic rearrangement leads to duplication of the *rcan-1* gene bodies, *rcan-1* shows lower expression. This finding gives many important insights into the role of gene duplication contributes to molecular evolution. It also directly provides the experimental clue that the genomic rearrangements generate gene duplication and result in fitness increase. Many potential interesting researches can be follow up: 1) *rcan-1* gene encodes calcirepressin which also plays a regulatory role in upstream of *crh-1* gene, the ortholog of human memory formation factor CREB[172,173]. The interesting follow-up research can study if *crh-1* pathway involved in fitness benefits of this complex genomic rearrangements. 2) Which biological traits does this complex genomic rearrangement regulate fitness? By designing the potential competition experiment using fluorescence-labeled animals, the followers can learn the distributions of the two competition strains. In addition, it could be

also intriguing to study if pheromones play important role in fitness benefits of the complex rearrangement. 3) At last but not least, how this *rcan-1* rearrangement occurred? Does it happened sequentially in multiple generations or it was generated during one DNA replication event? This complex DNA structure was generated in only 10 generations within the period building the N2*/LSJ RILs. To investigate how the DNA replication fork generated this complex rearrangement structure at ~4Kbp *rcan-1* promoter and gene body region? The potential hypothesis could be DNA replication fork encounter the stress surrounding *rcan-1* region and lead to the fork stalling and result in long single strand DNA region. This may lead to misannealing of the Okazaki fragment. The following homologous recombination repair events may potentially lead to this kind of complex genomic rearrangements. This can give scientists who studied genome stability maintenance and cancer biology more insights into the generation of complex genomic rearrangements.

In Chapter 4, I mainly quantify the fitness effect of the genetic variants in *C. elegans* under different feeding conditions and environmental factors. I identified multiple QTL show GxG and GxE that regulate animal's fitness. This part of work demonstrates that the experimental evolution approach can be used to systematically study GxE. Some interesting works could be followed up: 1) Comparing normal growth and starve growth condition, LSJ2 alleles including *Y61A9LA.3* at ChrV QTL shows GxG with *nurf-1* and GxE to regulate fitness. The exact function of this gene is unknown. It will be intriguing to do further study to identify its functional role in regulating fitness and how it interacts with *nurf-1* at molecular level. 2) It will be interesting to do further research to identify the adaptive mutations in LSJ2. The result suggests that multiple derived alleles across the

genome in LSJ2 contribute to the fitness effect. The potential project could apply CRISPR/Cas9 to make new ARLs or make new NILs that carry the LSJ2 alleles with interest to confirm their fitness effect.

My doctoral thesis research demonstrates that the experimental evolution approach is a powerful tool to identify the adaptive mutations in metazoan. The developed competition experiment strategy also shows a potential to performed in high-throughput level with automating platform like liquid-handler. The modified competition experiment could be further applied in the biomedical researches including identifying the mutations of chemoresistance/drug-resistance in tumor clones or pathogenic organisms.

APPENDIX

Table 2 - Genomic variants found in RIL_{hf}

CHROM	POS	REF	ALT	N2*(CX12311)	LSJ2	RILhf(CX12348)	PTM413 (rcan-1 NIL1)	PTM414 (rcan-1 NIL2)
I	3990614	C	T	0/0	0/0	0/0	0/1	0/1
I	11447094	G	C	0/0	0/0	0/0	1/1	1/1
I	13663159	A	T	0/0	0/0	0/0	1/1	1/1
II	240	C	T	0/0	0/0	0/0	0/1	0/1
II	291	A	G	0/0	0/0	0/0	0/1	0/1
II	5420981	T	G	0/0	0/0	0/0	1/1	0/0
II	9577378	T	A	0/0	0/0	0/0	1/1	1/1
II	12092091	C	G	0/0	0/0	0/0	0/0	1/1
II	14338974	C	T	0/0	0/0	0/0	0/1	0/1
III	985677	G	T	0/0	0/0	0/0	1/1	1/1
III	1014090	T	G	0/0	0/0	0/0	0/1	0/1
III	1890107	C	A	0/0	0/0	0/0	1/1	1/1
IV	1095706	C	A	0/0	0/0	0/0	0/0	1/1
IV	2828676	T	C	0/0	0/0	0/0	0/1	0/1
IV	3868771	G	A	0/0	0/0	0/0	1/1	0/0
IV	6519915	AT	TA	0/0	0/0	0/0	0/0	1/1
IV	7340720	T	G	0/0	0/0	0/0	1/1	0/0
IV	8700437	T	G	0/0	0/0	0/0	0/0	1/1
IV	8964224	A	G	0/0	0/0	0/0	1/1	1/1
IV	12319293	G	T	0/0	0/0	1/1	1/1	0/1
IV	17270018	C	T	0/0	0/0	0/0	1/1	1/1
V	10514873	A	T	0/0	0/0	0/0	0/0	1/1
V	12510124	CCT	TCA	0/0	0/0	0/0	0/1	0/1
V	17589629	T	C	0/0	0/0	0/0	0/1	0/0
X	1572624	G	C	0/0	0/0	0/0	1/1	1/1
X	3718036	G	A	0/0	0/0	0/0	0/0	1/1
X	4771953	T	A	0/0	0/0	0/0	1/1	0/0
X	10220078	A	C	0/0	0/0	0/0	1/1	0/0
X	11055944	C	A	0/0	0/0	0/0	0/0	1/1
X	12447310	T	C	0/0	0/0	0/0	1/1	1/1
X	15830406	G	A	0/0	0/0	0/0	1/1	1/1

Note: Reference Genome WS245; 0/0 represents homozygotes with REF(reference allele); 0/1 represents heterozygotes with REF(reference allele) and ALT(variant allele); 1/1 represents homozygotes with ALT(variant allele).

REFERENCES

1 Gibson, G. Decanalization and the origin of complex disease. *Nature Reviews Genetics* **10**, 134 (2009).

2 Cordell, H. J. Detecting gene–gene interactions that underlie human diseases. *Nature Reviews Genetics* **10**, 392 (2009).

3 Hunter, D. J. Gene–environment interactions in human diseases. *Nature Reviews Genetics* **6**, 287 (2005).

4 Eichler, E. E. *et al.* Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Reviews Genetics* **11**, 446-450 (2010).

5 Chang, A. J., Chronis, N., Karow, D. S., Marletta, M. A. & Bargmann, C. I. A distributed chemosensory circuit for oxygen preference in C. elegans. *PLoS Biol* **4**, e274, doi:10.1371/journal.pbio.0040274 (2006).

6 Mackay, T. F. Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nature Reviews Genetics* **15**, 22-33 (2014).

7 Anholt, R. R. & Mackay, T. F. Quantitative genetic analyses of complex behaviours in Drosophila. *Nature Reviews Genetics* **5**, 838-849 (2004).

8 Campbell, R. F., McGrath, P. T. & Paaby, A. B. Analysis of Epistasis in Natural Traits Using Model Organisms. *Trends in Genetics* (2018).

9 Baryshnikova, A. *et al.* Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nature methods* **7**, 1017 (2010).

10 Collins, S. R., Roguev, A. & Krogan, N. J. in *Methods in enzymology* Vol. 470 205-231 (Elsevier, 2010).

11 Costanzo, M. *et al.* The genetic landscape of a cell. *science* **327**, 425-431 (2010).

12 Lovejoy, C. A. *et al.* Functional genomic screens identify CINP as a genome maintenance protein. *Proceedings of the National Academy of Sciences* **106**, 19304-19309 (2009).

13 Breslow, D. K. *et al.* A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nature methods* **5**, 711 (2008).

14 Anholt, R. R. & Mackay, T. F. *Principles of behavioral genetics*. (Academic Press, 2009).

15 Broman, K. W. & Sen, S. *A Guide to QTL Mapping with R/qtl*. Vol. 46 (Springer, 2009).

16 Gou, L., Bloom, J. S. & Kruglyak, L. The genetic basis of mutation rate variation in yeast. *bioRxiv*, 338723 (2018).

17 Demogines, A., Smith, E., Kruglyak, L. & Alani, E. Identification and dissection of a complex DNA repair sensitivity phenotype in Baker's yeast. *PLoS genetics* **4**, e1000123 (2008).

18 Greene, J. S. *et al.* Balancing selection shapes density-dependent foraging behaviour. *Nature* **539**, 254-258, doi:10.1038/nature19848 (2016).

19 McGrath, P. T. *et al.* Quantitative mapping of a digenic behavioral trait implicates globin variation in C. elegans sensory behaviors. *Neuron* **61**, 692-699, doi:10.1016/j.neuron.2009.02.012 (2009).

20 Ghosh, R., Andersen, E. C., Shapiro, J. A., Gerke, J. P. & Kruglyak, L. Natural variation in a chloride channel subunit confers avermectin resistance in C. elegans. *Science* **335**, 574-578 (2012).

21 Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *The American Journal of Human Genetics* **90**, 7-24 (2012).

22 Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature genetics* **44**, 981 (2012).

23 Gerken, A. R., Eller, O. C., Hahn, D. A. & Morgan, T. J. Constraints, independence, and evolution of thermal plasticity: probing genetic architecture of long-and short-term thermal acclimation. *Proceedings of the National Academy of Sciences*, 201503456 (2015).

24 Fournier-Level, A. *et al.* A map of local adaptation in Arabidopsis thaliana. *Science* **334**, 86-89 (2011).

25 Cook, D. E., Zdraljevic, S., Roberts, J. P. & Andersen, E. C. CeNDR, the Caenorhabditis elegans natural diversity resource. *Nucleic acids research* **45**, D650-D657 (2016).

26 Evans, K. S. *et al.* Correlations of genotype with climate parameters suggest Caenorhabditis elegans niche adaptations. *G3: Genes, Genomes, Genetics* **7**, 289-298 (2017).

27 Klein, H. *et al.* Bulked-Segregant Analysis Coupled to Whole Genome Sequencing (BSA-Seq) for Rapid Gene Cloning in Maize. *G3: Genes, Genomes, Genetics* **8**, 3583-3592 (2018).

28 Parts, L. *et al.* Revealing the genetic structure of a trait by sequencing a population under selection. *Genome research*, gr. 116731.116110 (2011).

29 Zou, C., Wang, P. & Xu, Y. Bulked sample analysis in genetics, genomics and crop improvement. *Plant biotechnology journal* **14**, 1941-1955 (2016).

30 Ehrenreich, I. M. *et al.* Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* **464**, 1039 (2010).

31 Pool, J. E. Genetic mapping by bulk segregant analysis in Drosophila: experimental design and simulation-based inference. *Genetics* **204**, 1295-1306 (2016).

32 Burga, A., Ben-David, E., Vergara, T. L., Boocock, J. & Kruglyak, L. Fast genetic mapping of complex traits in C. elegans using millions of individuals in bulk. *BioRxiv*, 428870 (2018).

33 Garland, T. & Rose, M. R. *Experimental evolution: concepts, methods, and applications of selection experiments*. (University of California Press Berkeley, 2009).

34 Shipley, A. (JSTOR, 1910).

35 Blount, Z. D., Lenski, R. E. & Losos, J. B. Contingency and determinism in evolution: Replaying life's tape. *Science* **362**, eaam5979 (2018).

36 Barrick, J. E. *et al.* Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature* **461**, 1243 (2009).

37 Woods, R. J. *et al.* Second-order selection for evolvability in a large Escherichia coli population. *Science* **331**, 1433-1436 (2011).

38 Barrick, J. E. & Lenski, R. E. Genome dynamics during experimental evolution. *Nature Reviews Genetics* **14**, 827 (2013).

39 Maddamsetti, R., Lenski, R. E. & Barrick, J. E. Adaptation, clonal interference, and frequency-dependent interactions in a long-term evolution experiment with Escherichia coli. *Genetics*, genetics. 115.176677 (2015).

40 Blount, Z. D., Barrick, J. E., Davidson, C. J. & Lenski, R. E. Genomic analysis of a key innovation in an experimental Escherichia coli population. *Nature* **489**, 513 (2012).

41 Lang, G. I. *et al.* Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* **500**, 571 (2013).

42 Kryazhimskiy, S., Rice, D. P., Jerison, E. R. & Desai, M. M. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* **344**, 1519-1522 (2014).

43 Levy, S. F. *et al.* Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* **519**, 181 (2015).

44 Venkataram, S. *et al.* Development of a comprehensive genotype-to-fitness map of adaptation-driving mutations in yeast. *Cell* **166**, 1585-1596. e1522 (2016).

45 Kinnersley, M. *et al.* Evolutionary dynamics of de novo mutations and mutant lineages arising in a simple, constant environment. *BioRxiv*, 540625 (2019).

46 Kacar, B., Ge, X., Sanyal, S. & Gaucher, E. A. Experimental evolution of Escherichia coli harboring an ancient translation protein. *Journal of molecular evolution* **84**, 69-84 (2017).

47 Randall, R. N., Radford, C. E., Roof, K. A., Natarajan, D. K. & Gaucher, E. A. An experimental phylogeny to benchmark ancestral sequence reconstruction. *Nature communications* **7**, 12847 (2016).

48 Ratcliff, W. C. *et al.* Experimental evolution of an alternating uni-and multicellular life cycle in Chlamydomonas reinhardtii. *Nature communications* **4**, 2742 (2013).

49 Ratcliff, W. C., Fankhauser, J. D., Rogers, D. W., Greig, D. & Travisano, M. Origins of multicellular evolvability in snowflake yeast. *Nature communications* **6**, 6102 (2015).

50 Denver, D. R. *et al.* Selective sweeps and parallel mutation in the adaptive recovery from deleterious mutation in Caenorhabditis elegans. *Genome research* **20**, 1663-1671 (2010).

51 Farslow, J. C. *et al.* Rapid Increase in frequency of gene copy-number variants during experimental evolution in Caenorhabditis elegans. *BMC genomics* **16**, 1044 (2015).

52 Chelo, I. M., Nédli, J., Gordo, I. & Teotónio, H. An experimental test on the probability of extinction of new genetic variants. *Nature communications* **4** (2013).

53 Greene, J. S., Dobosiewicz, M., Butcher, R. A., McGrath, P. T. & Bargmann, C. I. Regulatory changes in two chemoreceptor genes contribute to a Caenorhabditis elegans QTL for foraging behavior. *Elife* **5** (2016).

54 Morran, L. T., Parmenter, M. D. & Phillips, P. C. Mutation load and rapid adaptation favour outcrossing over self-fertilization. *Nature* **462**, 350-352 (2009).

55 Morran, L. T., Cappy, B. J., Anderson, J. L. & Phillips, P. C. SEXUAL PARTNERS FOR THE STRESSED: FACULTATIVE OUTCROSSING IN THE SELF ‐ FERTILIZING NEMATODE CAENORHABDITIS ELEGANS. *Evolution* **63**, 1473-1482 (2009).

56 Anderson, J. L., Reynolds, R. M., Morran, L. T., Tolman-Thompson, J. & Phillips, P. C. Experimental evolution reveals antagonistic pleiotropy in reproductive timing but not life span in Caenorhabditis elegans. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* **66**, 1300-1308 (2011).

57 Carvalho, S., Phillips, P. C. & Teotónio, H. Hermaphrodite life history and the maintenance of partial selfing in experimental populations of Caenorhabditis elegans. *BMC evolutionary biology* **14**, 117 (2014).

58 Sikkink, K. L., Ituarte, C. M., Reynolds, R. M., Cresko, W. A. & Phillips, P. C. The transgenerational effects of heat stress in the nematode Caenorhabditis remanei are negative and rapidly eliminated under direct selection for increased stress resistance in larvae. *Genomics* **104**, 438-446 (2014).

59 Castillo, D. M., Burger, M. K., Lively, C. M. & Delph, L. F. Experimental evolution: Assortative mating and sexual selection, independent of local adaptation, lead to reproductive isolation in the nematode Caenorhabditis remanei. *Evolution* **69**, 3141-3155 (2015).

60 Morran, L. T., Schmidt, O. G., Gelarden, I. A., Parrish, R. C. & Lively, C. M. Running with the Red Queen: host-parasite coevolution selects for biparental sex. *Science* **333**, 216-218 (2011).

61 Robertson, R. M. & Margarida, M. *Methuselah flies: a case study in the evolution of aging*. (World Scientific, 2004).

62 Burke, M. K. *et al.* Genome-wide analysis of a long-term evolution experiment with Drosophila. *Nature* **467**, 587 (2010).

63 Turner, T. L., Stewart, A. D., Fields, A. T., Rice, W. R. & Tarone, A. M. Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in Drosophila melanogaster. *PLoS genetics* **7**, e1001336 (2011).

64 Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90 (2011).

65 Wang, Y. *et al.* Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* **512**, 155 (2014).

66 Kim, C. *et al.* Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. *Cell* **173**, 879-893. e813 (2018).

67 Casasent, A. K. *et al.* Multiclonal invasion in breast tumors identified by topographic single cell sequencing. *Cell* **172**, 205-217. e212 (2018).

68 Faltas, B. M. *et al.* Clonal evolution of chemotherapy-resistant urothelial carcinoma. *Nature genetics* **48**, 1490 (2016).

69 Xue, K. S. *et al.* Parallel evolution of influenza across multiple spatiotemporal scales. *Elife* **6**, e26875 (2017).

70 Sterken, M. G., Snoek, L. B., Kammenga, J. E. & Andersen, E. C. The laboratory domestication of Caenorhabditis elegans. *Trends Genet* **31**, 224-231, doi:10.1016/j.tig.2015.02.009 (2015).

71 McGrath, P. T. *et al.* Parallel evolution of domesticated Caenorhabditis species targets pheromone receptor genes. *Nature* **477**, 321-325, doi:10.1038/nature10378 (2011).

72 Goldstein, B. Sydney Brenner on the Genetics of Caenorhabditis elegans. *Genetics* **204**, 1-2, doi:10.1534/genetics.116.194084 (2016).

73 Brenner, S. The genetics of Caenorhabditis elegans. *Genetics* **77**, 71-94 (1974).

74 Zhao, Y. *et al.* Laboratory evolution from social to solitary behavior in the N2 reference strain is unnecessary for its fitness advantages. *bioRxiv*, 309997 (2018).

75 de Bono, M. & Bargmann, C. I. Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in C. elegans. *Cell* **94**, 679-689 (1998).

76 Persson, A. *et al.* Natural variation in a neural globin tunes oxygen sensing in wild Caenorhabditis elegans. *Nature* **458**, 1030-1033, doi:10.1038/nature07820 (2009).

77 Gray, J. M. *et al.* Oxygen sensation and social feeding mediated by a C. elegans guanylate cyclase homologue. *Nature* **430**, 317-322, doi:10.1038/nature02714 (2004).

78 Macosko, E. Z. *et al.* A hub-and-spoke circuit drives pheromone attraction and social behaviour in C. elegans. *Nature* **458**, 1171-1175, doi:10.1038/nature07886 (2009).

79 Rogers, C. *et al.* Inhibition of Caenorhabditis elegans social feeding by FMRFamide-related peptide activation of NPR-1. *Nature neuroscience* **6**, 1178 (2003).

80 Witham, E. *et al.* C. elegans Body Cavity Neurons Are Homeostatic Sensors that Integrate Fluctuations in Oxygen Availability and Internal Nutrient Reserves. *Cell Rep* **14**, 1641-1654, doi:10.1016/j.celrep.2016.01.052 (2016).

81 Duveau, F. & Felix, M. A. Role of pleiotropy in the evolution of a cryptic developmental variation in Caenorhabditis elegans. *PLoS Biol* **10**, e1001230, doi:10.1371/journal.pbio.1001230 (2012).

82 Large, E. E. *et al.* Selection on a subunit of the NURF chromatin remodeler modifies life history traits in a domesticated strain of Caenorhabditis elegans. *PLoS genetics* **12**, e1006219 (2016).

83 Large, E. E. *et al.* Modeling of a negative feedback mechanism explains antagonistic pleiotropy in reproduction in domesticated Caenorhabditis elegans strains. *PLoS genetics* **13**, e1006769 (2017).

84 Gould, S. J. & Lewontin, R. C. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc R Soc Lond B Biol Sci* **205**, 581-598 (1979).

85 Fisher, K. J. & Lang, G. I. Experimental evolution in fungi: An untapped resource. *Fungal Genet Biol* **94**, 88-94, doi:10.1016/j.fgb.2016.06.007 (2016).

86 Lenski, R. E. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *ISME J* **11**, 2181-2194, doi:10.1038/ismej.2017.69 (2017).

87 Teotonio, H., Estes, S., Phillips, P. C. & Baer, C. F. Experimental Evolution with Caenorhabditis Nematodes. *Genetics* **206**, 691-716, doi:10.1534/genetics.115.186288 (2017).

88 Alfred, J. & Baldwin, I. T. New opportunities at the wild frontier. *Elife* **4**, doi:10.7554/eLife.06956 (2015).

89 Gladfelter, A. S. How nontraditional model systems can save us. *Mol Biol Cell* **26**, 3687-3689, doi:10.1091/mbc.E15-06-0429 (2015).

90 Goldstein, B. & King, N. The Future of Cell Biology: Emerging Model Organisms. *Trends Cell Biol* **26**, 818-824, doi:10.1016/j.tcb.2016.08.005 (2016).

91 Russell, J. J. *et al.* Non-model model organisms. *BMC Biol* **15**, 55, doi:10.1186/s12915-017-0391-5 (2017).

92 Orozco-terWengel, P. *et al.* Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. *Mol Ecol* **21**, 4931-4941, doi:10.1111/j.1365-294X.2012.05673.x (2012).

93 Goto, T., Tanave, A., Moriwaki, K., Shiroishi, T. & Koide, T. Selection for reluctance to avoid humans during the domestication of mice. *Genes Brain Behav* **12**, 760-770, doi:10.1111/gbb.12088 (2013).

94 Kasahara, T., Abe, K., Mekada, K., Yoshiki, A. & Kato, T. Genetic variation of melatonin productivity in laboratory mice under domestication. *Proc Natl Acad Sci U S A* **107**, 6412-6417, doi:10.1073/pnas.0914399107 (2010).

95 Marks, M. E. *et al.* The genetic basis of laboratory adaptation in Caulobacter crescentus. *J Bacteriol* **192**, 3678-3688, doi:10.1128/JB.00255-10 (2010).

96 Stanley, C. E., Jr. & Kulathinal, R. J. Genomic signatures of domestication on neurogenetic genes in Drosophila melanogaster. *BMC Evol Biol* **16**, 6, doi:10.1186/s12862-015-0580-1 (2016).

97 Yvert, G. *et al.* Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. *Nat Genet* **35**, 57-64, doi:10.1038/ng1222 (2003).

98 Coates, J. C. & de Bono, M. Antagonistic pathways in neurons exposed to body fluid regulate social feeding in Caenorhabditis elegans. *Nature* **419**, 925-929, doi:10.1038/nature01170 (2002).

99 Chang, A. J. & Bargmann, C. I. Hypoxia and the HIF-1 transcriptional pathway reorganize a neuronal circuit for oxygen-dependent behavior in Caenorhabditis elegans. *Proc Natl Acad Sci U S A* **105**, 7321-7326, doi:10.1073/pnas.0802164105 (2008).

100 Andersen, E. C., Bloom, J. S., Gerke, J. P. & Kruglyak, L. A variant in the neuropeptide receptor npr-1 is a major determinant of Caenorhabditis elegans growth and physiology. *PLoS Genet* **10**, e1004156, doi:10.1371/journal.pgen.1004156 (2014).

101 Gloria-Soria, A. & Azevedo, R. B. npr-1 Regulates foraging and dispersal strategies in Caenorhabditis elegans. *Curr Biol* **18**, 1694-1699, doi:10.1016/j.cub.2008.09.043 (2008).

102 Noble, L. M. *et al.* Polygenicity and Epistasis Underlie Fitness-Proximal Traits in the Caenorhabditis elegans Multiparental Experimental Evolution (CeMEE) Panel. *Genetics* **207**, 1663-1685, doi:10.1534/genetics.117.300406 (2017).

103 Weber, K. P. *et al.* Whole genome sequencing highlights genetic changes associated with laboratory domestication of C. elegans. *PLoS One* **5**, e13922, doi:10.1371/journal.pone.0013922 (2010).

104 Evans, K. S. *et al.* Correlations of Genotype with Climate Parameters Suggest Caenorhabditis elegans Niche Adaptations. *G3 (Bethesda)* **7**, 289-298, doi:10.1534/g3.116.035162 (2017).

105 Large, E. E. *et al.* Selection on a Subunit of the NURF Chromatin Remodeler Modifies Life History Traits in a Domesticated Strain of Caenorhabditis elegans. *PLoS Genet* **12**, e1006219, doi:10.1371/journal.pgen.1006219 (2016).

106 Arribere, J. A. *et al.* Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in Caenorhabditis elegans. *Genetics* **198**, 837-846, doi:10.1534/genetics.114.169730 (2014).

107 Bernstein, M. R. & Rockman, M. V. Fine-Scale Crossover Rate Variation on the Caenorhabditis elegans X Chromosome. *G3 (Bethesda)* **6**, 1767-1776, doi:10.1534/g3.116.028001 (2016).

108 Bendesky, A. *et al.* Long-range regulatory polymorphisms affecting a GABA receptor constitute a quantitative trait locus (QTL) for social behavior in Caenorhabditis elegans. *PLoS Genet* **8**, e1003157, doi:10.1371/journal.pgen.1003157 (2012).

109 Bersaglieri, T. *et al.* Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet* **74**, 1111-1120, doi:10.1086/421051 (2004).

110 Li, W. H. The first arrival time and mean age of a deleterious mutant gene in a finite population. *Am J Hum Genet* **27**, 274-286 (1975).

111 Cheung, B. H., Cohen, M., Rogers, C., Albayram, O. & de Bono, M. Experience-dependent modulation of C. elegans behavior by ambient oxygen. *Curr Biol* **15**, 905-917, doi:10.1016/j.cub.2005.04.017 (2005).

112 Stern, S., Kirst, C. & Bargmann, C. I. Neuromodulatory Control of Long-Term Behavioral Patterns and Individuality across Development. *Cell* **171**, 1649-1662 e1610, doi:10.1016/j.cell.2017.10.041 (2017).

113 Reddy, K. C., Andersen, E. C., Kruglyak, L. & Kim, D. H. A polymorphism in npr-1 is a behavioral determinant of pathogen susceptibility in C. elegans. *Science* **323**, 382-384, doi:10.1126/science.1166527 (2009).

114 Garigan, D. *et al.* Genetic analysis of tissue aging in Caenorhabditis elegans: a role for heat-shock factor and bacterial proliferation. *Genetics* **161**, 1101-1112 (2002).

115 Gems, D. & Riddle, D. L. Defining wild-type life span in Caenorhabditis elegans. *J Gerontol A Biol Sci Med Sci* **55**, B215-219 (2000).

116 Styer, K. L. *et al.* Innate immunity in Caenorhabditis elegans is regulated by neurons expressing NPR-1/GPCR. *Science* **322**, 460-464, doi:10.1126/science.1163673 (2008).

117 Zuckerman, B. *et al.* Characterization of gene expression associated with the adaptation of the nematode C. elegans to hypoxia and reoxygenation stress reveals an unexpected function of the neuroglobin GLB-5 in innate immunity. *Free Radic Biol Med* **108**, 858-873, doi:10.1016/j.freeradbiomed.2017.05.007 (2017).

118 Cook, D. E., Zdraljevic, S., Roberts, J. P. & Andersen, E. C. CeNDR, the Caenorhabditis elegans natural diversity resource. *Nucleic Acids Res* **45**, D650-D657, doi:10.1093/nar/gkw893 (2017).

119 Boeck, M. E. *et al.* The time-resolved transcriptome of C. elegans. *Genome Res* **26**, 1441-1450, doi:10.1101/gr.202663.115 (2016).

120 Hubbard, E. J. & Greenstein, D. Introduction to the germ line. *WormBook*, 1-4, doi:10.1895/wormbook.1.18.1 (2005).

121 Large, E. E. *et al.* Modeling of a negative feedback mechanism explains antagonistic pleiotropy in reproduction in domesticated Caenorhabditis elegans strains. *PLoS Genet* **13**, e1006769, doi:10.1371/journal.pgen.1006769 (2017).

122 Anderson, J. L., Reynolds, R. M., Morran, L. T., Tolman-Thompson, J. & Phillips, P. C. Experimental evolution reveals antagonistic pleiotropy in reproductive timing but not life span in Caenorhabditis elegans. *J Gerontol A Biol Sci Med Sci* **66**, 1300-1308, doi:10.1093/gerona/glr143 (2011).

123 Ando, R. *et al.* Feeding responses to several neuropeptide Y receptor agonists in the neonatal chick. *Eur J Pharmacol* **427**, 53-59 (2001).

124 Lecklin, A. *et al.* Receptor subtypes Y1 and Y5 mediate neuropeptide Y induced feeding in the guinea-pig. *Br J Pharmacol* **135**, 2029-2037, doi:10.1038/sj.bjp.0704667 (2002).

125 Matsuda, K. Recent advances in the regulation of feeding behavior by neuropeptides in fish. *Ann N Y Acad Sci* **1163**, 241-250, doi:10.1111/j.1749-6632.2008.03619.x (2009).

126 Gomez-Amaro, R. L. *et al.* Measuring Food Intake and Nutrient Absorption in Caenorhabditis elegans. *Genetics* **200**, 443-454, doi:10.1534/genetics.115.175851 (2015).

127 Fang-Yen, C., Avery, L. & Samuel, A. D. Two size-selective mechanisms specifically trap bacteria-sized food particles in Caenorhabditis elegans. *Proc Natl Acad Sci U S A* **106**, 20093-20096, doi:10.1073/pnas.0904036106 (2009).

128 Abergel, Z., Chatterjee, A. K., Zuckerman, B. & Gross, E. Regulation of Neuronal Oxygen Responses in C. elegans Is Mediated through Interactions between Globin 5 and the H-NOX Domains of Soluble Guanylate Cyclases. *J Neurosci* **36**, 963-978, doi:10.1523/JNEUROSCI.3170-15.2016 (2016).

129 Gross, E. *et al.* GLOBIN-5-dependent O2 responses are regulated by PDL-1/PrBP that targets prenylated soluble guanylate cyclases to dendritic endings. *J Neurosci* **34**, 16726-16738, doi:10.1523/JNEUROSCI.5368-13.2014 (2014).

130 Oda, S., Toyoshima, Y. & de Bono, M. Modulation of sensory information processing by a neuroglobin in Caenorhabditis elegans. *Proc Natl Acad Sci U S A* **114**, E4658-E4665, doi:10.1073/pnas.1614596114 (2017).

131 Laurent, P. *et al.* Decoding a neural circuit controlling global animal state in C. elegans. *Elife* **4**, doi:10.7554/eLife.04241 (2015).

132 Jang, H. *et al.* Dissection of neuronal gap junction circuits that regulate social behavior in Caenorhabditis elegans. *Proc Natl Acad Sci U S A* **114**, E1263-E1272, doi:10.1073/pnas.1621274114 (2017).

133 Butcher, R. A. Decoding chemical communication in nematodes. *Nat Prod Rep* **34**, 472-477, doi:10.1039/c7np00007c (2017).

134 Ludewig, A. H. & Schroeder, F. C. Ascaroside signaling in C. elegans. *WormBook*, 1-22, doi:10.1895/wormbook.1.155.1 (2013).

135 Fenk, L. A. & de Bono, M. Memory of recent oxygen experience switches pheromone valence in Caenorhabditis elegans. *Proc Natl Acad Sci U S A* **114**, 4195-4200, doi:10.1073/pnas.1618934114 (2017).

136 Jang, H. *et al.* Neuromodulatory state and sex specify alternative behaviors through antagonistic synaptic pathways in C. elegans. *Neuron* **75**, 585-592, doi:10.1016/j.neuron.2012.06.034 (2012).

137 Butcher, R. A. *et al.* Biosynthesis of the Caenorhabditis elegans dauer pheromone. *Proc Natl Acad Sci U S A* **106**, 1875-1879, doi:10.1073/pnas.0810338106 (2009).

138 Zhang, S. O. *et al.* Genetic and dietary regulation of lipid droplet expansion in Caenorhabditis elegans. *Proc Natl Acad Sci U S A* **107**, 4640-4645, doi:10.1073/pnas.0912308107 (2010).

139 Park, S. & Paik, Y. K. Genetic deficiency in neuronal peroxisomal fatty acid beta-oxidation causes the interruption of dauer development in Caenorhabditis elegans. *Sci Rep* **7**, 9358, doi:10.1038/s41598-017-10020-x (2017).

140 Joo, H. J. *et al.* Caenorhabditis elegans utilizes dauer pheromone biosynthesis to dispose of toxic peroxisomal fatty acids for cellular homoeostasis. *Biochem J* **422**, 61-71, doi:10.1042/BJ20090513 (2009).

141 Li, S. *et al.* A Genetic Screen for Mutants with Supersized Lipid Droplets in Caenorhabditis elegans. *G3 (Bethesda)* **6**, 2407-2419, doi:10.1534/g3.116.030866 (2016).

142 Hyun, M. *et al.* Fat Metabolism Regulates Satiety Behavior in C. elegans. *Sci Rep* **6**, 24841, doi:10.1038/srep24841 (2016).

143 Gibson, G. & Dworkin, I. Uncovering cryptic genetic variation. *Nat Rev Genet* **5**, 681-690, doi:10.1038/nrg1426 (2004).

144 White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode Caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci* **314**, 1-340 (1986).

145 Hussey, R. *et al.* Pheromone-sensing neurons regulate peripheral lipid metabolism in Caenorhabditis elegans. *PLoS Genet* **13**, e1006806, doi:10.1371/journal.pgen.1006806 (2017).

146 Varet, H., Brillet-Guéguen, L., Coppée, J.-Y. & Dillies, M.-A. SARTools: a DESeq2-and edgeR-based R pipeline for comprehensive differential analysis of RNA-Seq data. *PLoS One* **11**, e0157022 (2016).

147 Chen, Y., Lun, A. T. & Smyth, G. K. in *Statistical analysis of next generation sequencing data* 51-74 (Springer, 2014).

148 Zhuo, W., Lu, H. & McGrath, P. T. Microfluidic platform with spatiotemporally controlled micro-environment for studying long-term C. elegans developmental arrests. *Lab on a Chip* **17**, 1826-1833 (2017).

149 Stevens, R. Sex Chromosomes and Sex-linked Genes. *The Canadian Veterinary Journal* **10**, 277 (1969).

150 Taylor, J. S. & Raes, J. Duplication and divergence: the evolution of new genes and old ideas. *Annu Rev Genet* **38**, 615-643, doi:10.1146/annurev.genet.38.072902.092831 (2004).

151 Innan, H. & Kondrashov, F. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet* **11**, 97-108, doi:10.1038/nrg2689 (2010).

152 Fuentes, J. J. *et al.* DSCR1, overexpressed in Down syndrome, is an inhibitor of calcineurin-mediated signaling pathways. *Hum Mol Genet* **9**, 1681-1690 (2000).

153 Zhao, Y. *et al.* Changes to social feeding behaviors are not sufficient for fitness gains of the Caenorhabditis elegans N2 reference strain. *Elife* **7**, doi:10.7554/eLife.38675 (2018).

154 Flavell, S. W. *et al.* Serotonin and the neuropeptide PDF initiate and extend opposing behavioral states in C. elegans. *Cell* **154**, 1023-1035, doi:10.1016/j.cell.2013.08.001 (2013).

155 Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in C. elegans: extrachromosomal maintenance and integration of transforming sequences. *The EMBO journal* **10**, 3959-3970 (1991).

156 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics* **25**, 1754-1760 (2009).

157 Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

158 http://broadinstitute.github.io/picard/index.html.

159 Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907* (2012).

160 Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80-92 (2012).

161 Robinson, J. T. *et al.* Integrative genomics viewer. *Nature biotechnology* **29**, 24 (2011).

162 Lee, H. *et al.* A multi-channel device for high-density target-selective stimulation and long-term monitoring of cells and subcellular features in C. elegans. *Lab on a Chip* **14**, 4513-4522 (2014).

163 Evans, K. S. *et al.* Shared genomic regions underlie natural variation in diverse toxin responses. *Genetics* **210**, 1509-1525 (2018).

164 Ye, K., Cao, C., Lin, X., O'Brien, K. O. & Gu, Z. Natural selection on HFE in Asian populations contributes to enhanced non-heme iron absorption. *BMC genetics* **16**, 61 (2015).

165 Ye, K., Gao, F., Wang, D., Bar-Yosef, O. & Keinan, A. Dietary adaptation of FADS genes in Europe varied across time and geography. *Nature ecology & evolution* **1**, 0167 (2017).

166 Buckley, M. T. *et al.* Selection in Europeans on fatty acid desaturases associated with dietary changes. *Molecular biology and evolution* **34**, 1307-1318 (2017).

167 Bersaglieri, T. *et al.* Genetic signatures of strong recent positive selection at the lactase gene. *The American Journal of Human Genetics* **74**, 1111-1120 (2004).

168 Hahnel, S. R. *et al.* Extreme allelic heterogeneity at a Caenorhabditis elegans beta-tubulin locus explains natural resistance to benzimidazoles. *PLoS pathogens* **14**, e1007226 (2018).

169 Ruthenburg, A. J. *et al.* Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions. *Cell* **145**, 692-706 (2011).

170 Zhao, X. *et al.* BPTF promotes hepatocellular carcinoma growth by modulating hTERT signaling and cancer stem cell traits. *Redox biology* **20**, 427-441 (2019).

171 Tomlinson, G. A. & Rothstein, M. Nematode biochemistry: I. Culture methods. *Biochimica et biophysica acta* **63**, 465-470 (1962).

172 Li, W., Bell, H. W., Ahnn, J. & Lee, S.-K. Regulator of Calcineurin (RCAN-1) Regulates Thermotaxis Behavior in Caenorhabditiselegans. *Journal of molecular biology* **427**, 3457-3468 (2015).

173 Kim, S. S. & Seo, S. R. The regulator of calcineurin 1 (RCAN1/DSCR1) activates the cAMP response element-binding protein (CREB) pathway. *Journal of Biological Chemistry* **286**, 37841-37848 (2011).